



# Risk hull method for spectral regularization in linear statistical inverse problems

Clément Marteau

## ► To cite this version:

Clément Marteau. Risk hull method for spectral regularization in linear statistical inverse problems. ESAIM: Probability and Statistics, 2010, 14, pp.409-434. hal-00369284

**HAL Id: hal-00369284**

**<https://hal.science/hal-00369284>**

Submitted on 19 Mar 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Risk hull method for spectral regularization in linear statistical inverse problems

**Clément MARTEAU**

Université de Provence, CMI,  
39, rue F. Joliot-Curie,  
F-13 453 Marseille cedex 13, France.

## Abstract

We consider in this paper the statistical linear inverse problem  $Y = Af + \epsilon\xi$  where  $A$  denotes a compact operator,  $\epsilon$  a noise level and  $\xi$  a stochastic noise. The unknown function  $f$  has to be recovered from the indirect measurement  $Y$ . We are interested in the following approach: given a family of estimators, we want to select the best possible one. In this context, the unbiased risk estimation (URE) method is rather popular. Nevertheless, it is also very unstable. Recently, Cavalier and Golubev (2006) introduced the risk hull minimization (RHM) method. It significantly improves the performances of the standard URE procedure. However, it only concerns projection rules. Using recent developments on ordered processes, we prove in this paper that it can be extended to a large class of linear estimators.

## 1 Introduction

This paper is devoted to statistical linear inverse problems. We want to recover an unknown function  $f$  from noisy and indirect measurements. Formally, consider  $H$  and  $K$  two Hilbert spaces and  $A : H \rightarrow K$  a compact operator. We observe:

$$Y = Af + \epsilon\xi, \quad (1.1)$$

where  $f$  belongs to  $H$  and  $\epsilon\xi$  denotes some noise. This representation arises in many mathematical and physical domains. In the numerical literature, the noise  $\epsilon\xi$  is deterministic. For such a model, many recovering methods have been proposed. For a survey, we mention for instance [13], [17], [14] or [25]. In the statistical literature, one deals instead with stochastic perturbations. The Gaussian white noise model is the most used. The representation (1.1) is equivalent to:

$$\langle Y, g \rangle = \langle Af, g \rangle + \epsilon \langle \xi, g \rangle, \quad \forall g \in K, \quad (1.2)$$

where  $\langle \xi, g \rangle \sim \mathcal{N}(0, \|g\|^2)$  and  $\epsilon > 0$ . Given  $g_1, g_2 \in K$ ,  $\mathbb{E}\langle \xi, g_1 \rangle \langle \xi, g_2 \rangle = \langle g_1, g_2 \rangle$ . We refer to [18] for more details concerning the Gaussian white noise. We assume throughout this paper that the noise level  $\epsilon$  is known.

The singular value decomposition (SVD) setting provides a better understanding of the model (1.1). The operator  $A^*A$  is compact and selfadjoint. We denote by  $(b_k^2)_{k \in \mathbb{N}}$  the associated sequence of eigenvalues. The set of eigenfunctions  $(\phi_k)_{k \in \mathbb{N}}$  is assumed to be orthonormal. Then define  $\psi_k = b_k^{-1}A\phi_k$ ,  $\forall k \in \mathbb{N}$ . For all integer  $k$ , the following equality holds:

$$\begin{cases} A\phi_k = b_k\psi_k, \\ A^*\psi_k = b_k\phi_k. \end{cases} \quad (1.3)$$

The triple  $(b_k, \phi_k, \psi_k)_{k \in \mathbb{N}}$  is called the singular system for the operator  $A^*A$ . The associated representation matrix is diagonal. For all  $k \in \mathbb{N}$ , set  $g = \psi_k$  in (1.2) and use (1.3) in order to obtain the sequence space model:

$$y_k = b_k\theta_k + \epsilon\xi_k, \quad \forall k \in \mathbb{N}, \quad (1.4)$$

where  $\theta_k = \langle f, \phi_k \rangle$ . The  $\xi_k$  are i.i.d. standard Gaussian random variables since the set  $(\psi_k)_{k \in \mathbb{N}}$  is also orthonormal. In the  $L^2$  sense, the function  $f$  and  $\theta = (\theta_k)_{k \in \mathbb{N}}$  represent the same mathematical object. The sequence  $\theta$  has to be recovered from the noisy observations  $(y_k)_{k \in \mathbb{N}}$ . Since  $A^*A$  is compact,  $b_k \rightarrow 0$  as  $k \rightarrow +\infty$ . When  $k$  is large, the estimation of  $\theta_k$  is quite difficult: there is mainly noise in the observation  $y_k$ . The SVD can explicitly be computed for a large class of inverse problems, e.g. tomography (see [19]) or Biophotonic imaging (see [4]).

In some sense, the best approximate solution  $\hat{f}$  minimizes the distance between the observation and the image of the operator  $A$ . It corresponds to the least square solution and verifies  $\hat{f} = (A^*A)^{-1}A^*Y$ . Since  $A^*A$  is compact, it is not continuously invertible. The solution  $\hat{f}$  will not be convenient: it does not necessarily converge to  $f$  as  $\epsilon \rightarrow 0$ . In order to solve (1.1), one may use instead:

$$\hat{f}_t = \Phi_t(A^*A)A^*Y,$$

where  $\Phi_t$  approximates the function  $x \mapsto x^{-1}$  on the spectrum of  $A^*A$ . This is a regularization method. The parameter  $t$  is a regularization parameter. The performances of  $\hat{f}_t$  are related to the choice of  $t$  which is crucial for both numerical and statistical approaches. Some examples are presented in Sections 2.1 and 3.2.

In the SVD setting, a regularization method  $\Phi_t$  corresponds to a filter  $\lambda(t) = (\lambda_k(t))_{k \in \mathbb{N}}$ . It is a real sequence with values in  $[0, 1]$ . The associated estimator is:

$$\hat{f}_t = \sum_{k=1}^{+\infty} \lambda_k(t) b_k^{-1} y_k \phi_k.$$

The quality of  $\hat{f}_t$  is measured via its quadratic risk:

$$R(\theta, \lambda(t)) = \mathbb{E}_\theta \|\hat{f}_t - f\|^2 = \sum_{k=1}^{+\infty} (1 - \lambda_k(t))^2 \theta_k^2 + \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2(t) b_k^{-2}. \quad (1.5)$$

The behavior of  $R(\theta, \lambda(t))$  depends on both the function  $f$  and the regularization approach, i.e. the filter  $\lambda$  and the parameter  $t$ . For more details and some examples, we mention [15], [19], [11] or [12].

In this context, the unbiased risk estimation (URE) method is rather popular for choosing  $t$  since it does not require a priori informations on the solution  $f$ . It has been studied for instance in [8]. However, this method is rather unstable due to the ill-posedness of the problem. The risk hull minimization (RHM) method initiated in [9] proposes to take into account the variability of the problem through the construction of a risk hull. The definition is detailed below.

**Definition 1** *Let  $\Lambda$  a family of filters. A deterministic term  $V(\theta, \lambda)$  such that:*

$$\mathbb{E}_\theta \sup_{\lambda \in \Lambda} [\|\hat{f}_t - f\|^2 - V(\theta, \lambda)] \leq 0,$$

*is called a risk hull for  $\Lambda$ .*

The principle of the RHM scheme is to minimize the hull, through an appropriate estimator, instead of the quadratic risk. Specific phenomena are related to the Gaussian white noise model. They require particular approaches that may not be useful in numerical areas. The RHM method is a good example of such a procedure. It provides answers to specific problems of the statistical model (1.2) and improves the standard URE approach. The risk hull method has been developed for projection estimators. As proved in this paper, it can be extended to a large class of linear estimators. This class contains for instance Tikhonov estimators or Landweber iterative methods.

This paper is organized as follows. In Section 2, we present the RHM method and recall the main problems related to the choice of the regularization parameter  $t$  through some well-known properties of Tikhonov estimators. A risk hull for a wide family of spectral regularization schemes is constructed in Section 3. Section 4 contains the main results and the proofs are gathered in Section 5. Finally, Section 6 is devoted to ordered processes. A reader not familiar with this topic is advised to read this part before being interested in the proofs.

## 2 The risk hull principle

Consider the following problem: given a function  $f$  and a set of estimators  $\Lambda$ , we want to select the best possible one, i.e. that minimizes the quadratic risk. In most cases, the family  $\Lambda$  may be identified with a regularization operator  $\Phi_t$  indexed by  $t$  and a collection of parameters  $T$ . This collection may be finite or not. We want to select the parameter  $t$  in an adaptive way, i.e. without using some unknown information on the function  $f$ .

### 2.1 The Tikhonov estimator

In order to shed some light on this framework, we recall briefly some properties of the Tikhonov estimators. Obviously, these estimators are well-known. However, some steps may be useful in the sequel. For a survey on this procedure, both in the numerical and statistical domains, we mention for instance [14], [24], [6] or [5] for non-linear inverse problems.

The Tikhonov estimation is rather intuitive. One wants to control both the regularity of our estimator  $\tilde{f}$  and the fitting of  $A\tilde{f}$  with the data  $Y$ . Therefore, resolve the following optimization problem:

$$\tilde{f}_t = \arg \min_{f \in H} \{\|Af - Y\|^2 + t^{-1}\|f\|_a^2\}, \quad (2.1)$$

where  $t$  is a regularization parameter. For all  $f \in H$  and  $a \in \mathbb{R}$ ,  $\|f\|_a = \|(AA^*)^{-a}f\|$ . In some sense,  $(AA^*)^{-a}$  is a differential operator. The parameter  $a$  is chosen a priori. It is related to the expected regularity of the function  $f$ . In the literature, one often uses  $\alpha$  instead of  $t^{-1}$ . However, the notation (2.1) is consistent with Sections 3-6.

The solution of (2.1) is:

$$\tilde{f}_t = \Phi_t(A^*A)A^*Y, \text{ with } \Phi_t(A^*A) = (A^*A + t^{-1}(L^a)^*L^a)^{-1}, \text{ and } L^a = (AA^*)^{-a}. \quad (2.2)$$

In the SVD representation, the Tikhonov estimator becomes:

$$\tilde{f}_t = \sum_{k=1}^{+\infty} \lambda_k(t) b_k^{-1} y_k \phi_k, \text{ with } \lambda_k(t) = \frac{1}{1 + t^{-1}(b_k^{-2})^{1+2a}}, \quad \forall k \in \mathbb{N}. \quad (2.3)$$

The choice of  $t$  is crucial and has a real impact on the performances of  $\tilde{f}_t$ . It is a trade-off between the two sums in the right hand side of (1.5). Assume for instance that the function  $f$  belongs to a ball  $H_Q^\mu$  of radius  $Q$ :

$$f \in H_Q^\mu = \{g \in H : \|(A^*A)^{-\mu}g\|^2 \leq Q\} \Leftrightarrow \theta \in \Theta(s, Q) = \left\{ \vartheta : \sum_{k=1}^{+\infty} b_k^{-4\mu} \vartheta_k^2 \leq Q \right\}, \quad (2.4)$$

for some  $\mu > 0$ . In what follows, we note  $b_1 \lesssim b_2$  when there exists  $C > 0$  such that  $b_1 \leq Cb_2$ . Using simple algebra:

$$\sum_{k=1}^{+\infty} (1 - \lambda_k(t))^2 \theta_k^2 \lesssim t^{-\frac{2\mu}{(1+2a)}}, \quad (2.5)$$

provided  $\mu < 1 + 2a$ . For  $\mu$  larger than  $1 + 2a$ ,  $g_\alpha$  cannot be optimally bounded. The Tikhonov estimator does not attain the minimax rate of convergence on  $H_Q^\mu$ . It is said to be underqualified. For the special case  $a = 0$ , we obtain the well-known condition  $\mu < 1$ . The term  $1 + 2a$  is called qualification of the Tikhonov regularization (see [14] or [6] for more details).

Now consider the second sum. Assume that the sequence of eigenvalues possesses a polynomial behavior:  $(b_k)_{k \in \mathbb{N}} \sim (k^{-\beta})_{k \in \mathbb{N}}$  for some  $\beta > 0$ . The problem is said to be mildly ill-posed. Let  $n = n_t \in \mathbb{N}$  which will be chosen later:

$$\epsilon^2 \sum_{k=1}^{+\infty} \lambda_k(t)^2 b_k^{-2} = \epsilon^2 \sum_{k=1}^{n_t} \lambda_k(t)^2 b_k^{-2} + \epsilon^2 \sum_{k > n_t} \lambda_k(t)^2 b_k^{-2} \lesssim \epsilon^2 n_t^{2\beta+1} + \epsilon^2 t^{-2} n_t^{-8\beta a - 2\beta + 1}.$$

Using simple algebra, we obtain the following bound:

$$\epsilon^2 \sum_{k=1}^{+\infty} \lambda_k(t)^2 b_k^{-2} \lesssim \epsilon^2 t^{\frac{2\beta+1}{2\beta(1+2a)}}, \text{ setting } n_t = t^{\frac{1}{2\beta(1+2a)}}. \quad (2.6)$$

Consider the parameter  $t_0$  that make the trade-off between (2.5) and (2.6). The associated estimator attains the minimax rate of convergence on  $H_Q^\mu$  for  $\mu < 1 + 2a$ . The parameter  $t_0$  can easily be constructed with the a priori knowledge of  $\mu$  and  $Q$ . If it is not the case, it is necessary to use a data-driven parameter choice rule.

## 2.2 The risk hull method

In a minimax sense, the optimal regularization parameter depends essentially on the regularity and on the norm of  $f$ . Since these informations are in most cases unknown, the choice of  $t$  is somewhat difficult.

Several data driven regularization methods have been proposed. The unbiased risk estimation procedure is rather popular. The principle is intuitive: given  $f$  and a set of estimators  $\Lambda$  (or equivalently a collection  $T$  of regularization parameters), we want to select the best possible one, i.e. that minimizes the quadratic risk  $R(\theta, \lambda(t))$  (see (1.5)). Since this quantity depends on the unknown function  $f$ , this estimator is not available. It is called the oracle. However, we can construct an estimator  $U(y, \lambda(t))$  of  $R(\theta, \lambda(t))$  and then, minimize  $U(y, \lambda(t))$  instead of  $R(\theta, \lambda(t))$ . A natural estimator for  $R(\theta, \lambda(t))$  is:

$$U(y, \lambda(t)) = \sum_{k=1}^{+\infty} \{\lambda_k^2(t) - 2\lambda_k(t)\} (b_k^{-2} y_k^2 - \epsilon^2 b_k^{-2}) + \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2(t) b_k^{-2}.$$

This method has been applied by [10] on the set of blockwise monotone estimator. In a model selection context, [8] dealt with finite families of estimators. Sharp oracle inequalities have been obtained. However, numerical simulations are somewhat disappointing. This has been illustrated in [9] for the projection regularization. The same phenomenon occurs for Tikhonov estimators. For ill-posed inverse problems, the URE algorithm selects large regularization parameters when the oracle is typically small.

The instability of the URE procedure has already been discussed in the literature. In the last decades, some authors were interested in the criterion  $U(y, \lambda(t)) + \text{pen}(t)$  where  $\text{pen}(t) \rightarrow +\infty$  as  $t \rightarrow +\infty$ . A penalty is introduced in the estimator of  $R(\theta, \lambda(t))$ . This penalty is chosen in order to control the variability of the problem. We expect that smaller parameters will be selected. Several penalizations have been proposed for both direct and inverse problems. We mention for instance [1], [3], [16] or [21].

Other approaches leading to oracle inequalities in a more general context have been proposed. For instance, concerning the Lepski balancing principle we may mention [2], [23] or [5].

The risk hull minimization method (also denoted RHM) initiated by [9] provides an interesting alternative. Consider the simple example of projection filter (also called spectral cut-off):  $\lambda_k(t) = \mathbf{1}_{\{k \leq t\}}$ , for all  $k \in \mathbb{N}$ . Denote by  $\hat{\theta}_t$  the related estimator. The loss is:

$$l(\theta, \lambda(t)) = \|\hat{\theta}_t - \theta\|^2 = \sum_{k>t} \theta_k^2 + \epsilon^2 \sum_{k=1}^t b_k^{-2} \xi_k^2.$$

Due to the ill-posedness of the problem, the variance of  $l(\theta, t)$  is very large and explodes with  $t$ . However, this variability is neglected in the URE procedure. Indeed, we deal with  $R(\theta, \lambda(t)) = \mathbb{E}_\theta l(\theta, \lambda(t))$ . In order to take account of this variability, one may look after a risk hull, i.e. a deterministic quantity  $V(\theta, \lambda(t))$  verifying:

$$\mathbb{E}_\theta \sup_t \{l(\theta, \lambda(t)) - V(\theta, \lambda(t))\} \leq 0.$$

Then, we estimate and minimize  $V(\theta, \lambda(t))$  instead of  $R(\theta, \lambda(t))$ . The challenge is to find an explicit hull. It should not be too large in order to provide a good quality of estimation although a small hull will not be sufficiently stable. The hull  $V(\theta, \lambda(t))$  of [9] can be explicitly constructed. It is easily computable using for instance Monte-Carlo approximations. The related performances are significantly better than the standard URE procedure, both from a theoretical and numerical point of view. Note that the risk hull method leads in fact to a penalized URE criterion where the penalty is explicitly computable.

However, the projection estimation is rather rough. There exist several regularization methods with better performances. We may mention for instance the Tikhonov estimators presented in this section. The generalization of the RHM algorithm to a wide family of linear estimators may produce interesting results but is not obvious. It requires in particular the control of the process:

$$\eta(t) = \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2(t) b_k^{-2} (\xi_k^2 - 1), \quad t \geq 0.$$

When dealing with projection procedures,  $\eta(t)$ ,  $t \geq 0$  can be controlled as a Wiener process. These objects have been intensively studied and many results are available. In this paper, the assumptions concerning the sequence  $(\lambda_k(t))_{k \in \mathbb{N}}$  are rather weak. In such a situation, we will see that the process  $\eta(t)$ ,  $t \geq 0$  is said to be ordered. Some interesting properties have been established in [20] and [7]. It presents a different behavior compared to the well-known Wiener process. The generalization of the RHM algorithm requires advanced probabilistic tools, gathered in Section 6.

### 3 Risk hull

#### 3.1 Assumptions and construction

Here and in the sequel, assume that  $\Lambda$  is a family of monotones filters indexed by a positive parameter  $t$ , i.e.  $\Lambda = (\lambda(t))_{t \in T}$ , where  $T \subseteq \mathbb{R}^+$ . Each filter  $\lambda(t)$  may be noted  $\lambda$  or identified with the regularization parameter  $t$ . The meaning will be clear following the context. The associated linear estimator is denoted by  $\hat{\theta}_t$ . For all  $k \in \mathbb{N}$ , the function  $t \rightarrow \lambda_k(t)$  is assumed to be monotone non-decreasing. We require some additional assumptions on the family  $\Lambda$ .

**Polynomial hypotheses.** *There exist  $C_1, C_2, C_3, C_4$  positive constants and  $d > 0$  independent of  $t$  and  $\epsilon$  such that:*

$$\frac{\max_k \epsilon^2 \lambda_k^2(t) b_k^{-2}}{\sqrt{\sigma^2(t)}} \leq C_1 t^{-d/2}, \quad \text{where } \sigma^2(t) = \epsilon^4 \sum_{k=1}^{+\infty} \lambda_k^4(t) b_k^{-4}, \quad (3.1)$$

$$\exp \left[ -\frac{1}{4} \sum_{p=1}^{+\infty} \log \left( 1 + \frac{2s^2 \epsilon^4 \lambda_p^4(t) b_p^{-4}}{\sigma^2(t)} \right) \right] \leq \left( 1 + \frac{C_2 s^2}{t^d} \right)^{-t^d}, \quad \forall s \in \mathbb{R}, \quad (3.2)$$

and for all  $l \in \mathbb{N}$ ,

$$(C_3 t^d)^{-l/2+1} \leq \left| \sigma^2(t)^{-l/2} \epsilon^{2l} \sum_{p=1}^{+\infty} \lambda_p^{2l}(t) b_p^{-2l} \right| \leq (C_4 t^d)^{-l/2+1}. \quad (3.3)$$

In some sense, the inequalities (3.1)-(3.3) generalize the polynomial hypotheses of [9]. Some examples of regularization method satisfying (3.1)-(3.3) are presented in Section 3.2.

For all  $\lambda \in \Lambda$ , i.e. for all  $t \in T$ , introduce:

$$U_0(t) = \inf \{ u > 0 : \mathbb{E} \eta(t) \mathbf{1}_{\{\eta(t) > u\}} \leq \epsilon^2 \}, \quad \text{with } \eta(t) = \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2(t) b_k^{-2} (\xi_k^2 - 1). \quad (3.4)$$

The term  $\sigma^2(t)$  denotes the variance of the process  $\eta(t)$ ,  $t \geq 0$  up to some constant. This process is ordered: see Section 6 for more details. In particular, the function  $\sigma^2$  is monotone non-decreasing. The following lemma provides a lower bound for  $U_0(t)$ . This quantity will be useful in the following.

**Lemma 1** Assume that (3.1)-(3.3) hold. There exists  $T_0 > 0$  independent of  $\epsilon$  such that for all  $t > T_0$ ,

$$U_0(\lambda(t)) = U_0(t) \geq \sqrt{2\sigma^2(t) \log \left( \frac{\sigma^2(t)}{2\pi\epsilon^4} \right)},$$

where  $\sigma^2(t)$  is defined in (3.1).

A proof of this lemma is available in Section 5. We are now able to propose a risk hull. First introduce:

$$\lambda(t_0) = \lambda^0 = \arg \inf_{\lambda \in \Lambda} R(\theta, \lambda), \quad (3.5)$$

where  $R(\theta, \lambda)$  is the quadratic risk defined in (1.5). Hence,  $\lambda^0$  corresponds to the oracle filter for the family  $\Lambda$ . For all  $x \in \mathbb{R}^+$ , define also:

$$\omega(x) = \sup_{t \in T} \sup_k \lambda_k^2(t) b_k^{-2} \mathbf{1}_{\{\sum_{l=1}^{+\infty} \lambda_l^2(t) b_l^{-2} \leq x \sup_l \lambda_l^2(t) b_l^{-2}\}}, \quad (3.6)$$

and,

$$L_T = \log^2 S, \text{ with } S = \frac{\max_{t \in T} \sup_k b_k^{-2} \lambda_k^2(t)}{\min_{t \in T} \sup_k b_k^{-2} \lambda_k^2(t)}. \quad (3.7)$$

These terms have been introduced in [8]. The function  $\omega$  is explicitly computed in Section 3.2 for some particular families of estimators. The quantity  $S$  provides an information on the homogeneity of  $T$ . For  $S < +\infty$ , the collection  $T$  may be identified with an interval  $[t_1, t_{\max}]$  and  $S$  is linked to the ratio  $t_{\max}/t_1$ .

**Theorem 1** Assume that (3.1)-(3.3) hold. Then, there exist  $C^*, \bar{C} > 0$  such that, for all  $\theta$  and  $\alpha, \gamma > 0$ ,

$$\begin{aligned} V_\alpha(\theta, \lambda(t)) &= (1 + \gamma) \left\{ \sum_{k=1}^{+\infty} (1 - \lambda_k(t))^2 \theta_k^2 + \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2(t) b_k^{-2} + (1 + \alpha) U_0(\lambda(t)) \right\} \\ &\quad + \frac{C^* \epsilon^2}{\alpha} + \bar{C} \epsilon^2 \gamma^{-1} L_T \omega(\gamma^{-2} L_T) + \gamma R(\theta, \lambda(t_0)), \end{aligned} \quad (3.8)$$

is a risk hull, i.e.:

$$\mathbb{E}_\theta \sup_{t \in T} \left\{ \|\hat{\theta}_t - \theta\|^2 - V_\alpha(\theta, \lambda(t)) \right\} \leq 0.$$

The hull of [9] is constructed for projection estimators. Here, the main difference is contained in the residual term  $\bar{C} \epsilon^2 \gamma^{-1} L_T \omega(\gamma^{-2} L_T)$ . The hull is somewhat less precise. This can be explained by the structure of the stochastic processes involved in the loss. Indeed, when considering projection estimators, one essentially deals with Wiener processes. These processes are well-known and may be easily controlled. In particular:

$$P \left( \max_{t \geq 0} [W(t) - \gamma t] \geq x \right) \leq \exp(-2\gamma x), \quad \forall \gamma > 0 \text{ and } x \in \mathbb{R}^+,$$

where  $W(t), t \geq 0$  denotes a Wiener process. The proof of Theorem 1 is based on the theory of ordered processes summarized in Section 6 (see also [20] or [7]). These objects are well understood but more difficult to control. At best, it is (at the moment) possible to prove that for all  $p \in \mathbb{N}$ :

$$P \left( \max_{t \geq 0} [\rho(t) - \gamma \Sigma^2(t)] \geq x \right) \leq \frac{C p^p}{(\gamma x)^p}, \quad \forall \gamma > 0 \text{ and } x \in \mathbb{R}^+, \quad (3.9)$$

for a given ordered process  $\rho(t)$ ,  $t \geq 0$  of variance  $\Sigma^2(t)$ . Here,  $C$  denotes a positive constant independent of  $x$ . In light of the proofs of Theorem 3.2, it seems difficult to improve (3.9) in order to obtain an exponential inequality as for the Wiener process.

### 3.2 Examples

In this section, we illustrate the polynomial hypotheses (3.1)-(3.3). Assume that the sequence  $(b_k)_{k \in \mathbb{N}}$  possesses a polynomial behavior, i.e.  $(b_k)_{k \in \mathbb{N}} \sim (k^{-\beta})_{k \in \mathbb{N}}$  for some positive parameter  $\beta$ . The degree of ill-posedness  $\beta$  is assumed to be known.

We consider three different procedures: projection, Tikhonov and Landweber iterative methods.

EXAMPLE 1. (Projection estimators)

Obviously, Assumption (3.1)-(3.3) are satisfied for projection filters since the aim of this paper is to generalize [9]. Consider the set:

$$\Lambda_{\text{Proj}} = \left\{ \lambda = (\lambda_k(t))_{k \in \mathbb{N}} : \forall k \in \mathbb{N}, \lambda_k = \mathbf{1}_{\{k \leq t\}}, t \in \mathbb{N} \right\}.$$

The projection estimator is also called spectral cut-off in the literature.

Assumption (3.2) is verified in the proof of Lemma 1 of [9] with  $d = 1$ . Since  $(b_k)_{k \in \mathbb{N}} \sim (k^{-\beta})_{k \in \mathbb{N}}$ :

$$\sigma^2(t) = \sum_{k=1}^t b_k^{-4} \simeq t^{4\beta+1}, \text{ and } \sum_{p=1}^{+\infty} \lambda_p^{2l}(t) b_p^{-2l} = \sum_{p=1}^t b_p^{-2l} \simeq t^{2\beta l+1}, \forall l \in \mathbb{N},$$

uniformly in  $t \in \mathbb{N}$ . Therefore, (3.3) holds with  $d = 1$ . Then, remark that:

$$\frac{\max_k \lambda_k^2(t) b_k^{-2}}{\sqrt{\sigma^2(t)}} \simeq \frac{t^{2\beta}}{t^{2\beta+1/2}} = t^{-1/2}.$$

[8] proved that  $\omega(x) \simeq x^{2\beta}$  for all  $x \in \mathbb{R}^+$  and for some  $C > 0$ . This concludes Example 1.

EXAMPLE 2. (Tikhonov estimators)

We use the same notations of Section 2.1. Consider the family:

$$\Lambda_{\text{Tikh}} = \left\{ \lambda = (\lambda_k(t))_{k \in \mathbb{N}} : \forall k \in \mathbb{N}, \lambda_k = \frac{1}{1 + t^{-1}(b_k^{-2})^{(1+2a)}}, t \in \mathbb{R}^+ \right\}.$$

It is possible to show:

$$\sigma^2(t) = \epsilon^4 \sum_{k=1}^{+\infty} \lambda_k^4(t) b_k^{-4} \simeq \epsilon^4 (t^d)^{4\beta+1} \text{ and } \sum_{p=1}^{+\infty} \lambda_p^l(t) b_p^{-l} \simeq (t^d)^{2\beta l+1}, \forall l \in \mathbb{N}.$$

with  $d = 1/(2\beta(1+2a))$  (see (2.6)). Hence, (3.3) holds. For all  $t$ , let  $n_t$  defined in (2.6). For all  $k \in \{n_t/2, \dots, n_t\}$ :

$$\lambda_k^2 b_k^{-2} = \frac{b_k^{-2}}{[1 + t^{-1}(b_k^{-2})^{(1+a)}]^2} \geq C k^{2\beta} \geq C \frac{n_t^{2\beta}}{2^{2\beta}}.$$

Therefore, for all  $t \in \mathbb{R}^+$  and  $s \in \mathbb{R}$ ,

$$\begin{aligned} \exp \left[ -\frac{1}{4} \sum_{l=1}^{+\infty} \log \left( 1 + \frac{2s^2 \epsilon^4 \lambda_l^4(t) b_l^{-4}}{\sigma^2(t)} \right) \right] &\leq \exp \left[ -\frac{1}{4} \sum_{l=n_t/2}^{n_t} \log \left( 1 + \frac{2s^2 \epsilon^4 \lambda_l^4(t) b_l^{-4}}{\sigma^2(t)} \right) \right], \\ &\leq \exp \left[ -\frac{1}{4} \sum_{l=n_t/2}^{n_t} \log \left( 1 + \frac{C s^2 n_t^{4\beta}}{n_t^{4\beta+1}} \right) \right], \\ &\leq \left( 1 + \frac{C_1 s^2}{n_t} \right)^{-n_t/8} = \left( 1 + \frac{C_1 s^2}{t^d} \right)^{-t^d/8}. \end{aligned} \quad (3.10)$$

This prove (3.2). Finally, remark that:

$$\frac{\max_k \lambda_k^2(t) b_k^{-2}}{\sqrt{\sigma^2(t)}} \simeq \frac{\lambda_{n_t}^2(t) b_{n_t}^{-2}}{\sqrt{\sigma^2(t)}} \simeq \frac{n_t^{2\beta}}{n_t^{2\beta+1/2}} = t^{-d/2}. \quad (3.11)$$



Moreover, we can prove, using for instance (3.11), that  $\omega(x) \leq Cx^{2\beta}$  for all  $x \in \mathbb{R}^+$  and for some  $C > 0$ .

EXAMPLE 3. (Iterated Tikhonov regularization)

This method presents an interesting alternative to the previous one. The qualification of the standard Tikhonov estimator is 1 if we set  $a = 0$  in (2.1). This can be enhanced using iterations. In the following, we refer to  $\hat{f}_j$  as the iterated Tikhonov estimator of order  $j$ . Set  $\hat{f}_0 = 0$ . Given  $j \in \mathbb{N}$  and  $\hat{f}_{j-1}$ ,  $\hat{f}_j$  is defined as the unique solution of:

$$(A^*A + t^{-1}I)\hat{f}_j = A^*T + t^{-1}\hat{f}_{j-1},$$

where  $I$  denotes the identity operator. In the SVD setting, the following representation arises:

$$\hat{f}_j = \sum_{k=1}^{+\infty} \lambda_k(t) b_k^{-1} y_k \phi_k, \text{ with } \lambda_k(t) = \frac{(t^{-1} + b_k^2)^j - t^{-j}}{(t^{-1} + b_k^2)^j}, \forall k \in \mathbb{N}.$$

The qualification corresponds to the number of iterations. It is possible to prove that  $\max_k \lambda_k^2(t) b_k^{-2} \lesssim n_t^{2\beta}$  and  $\sigma^2(t) \geq n_t^{4\beta+1}$ , with  $n_t = t^{-1/2\beta}$ . Hence, (3.1) holds. Then, for all  $k \in \{n_t/2, \dots, n_t\}$ ,

$$\lambda_k^2(t) b_k^{-2} \geq \left[ (1 + t^{-1} b_k^{-2})^j - t^{-j} b_k^{-2j} \right] b_k^{-2} \geq C n_t^{2\beta},$$

for some  $C > 0$ . Using the same algebra as in (3.10), one obtain (3.2). The proof of (3.3) and the bound of  $\omega$  essentially follows the same lines.

EXAMPLE 4. (Landweber iterative method)

This procedure is rather interesting from a numerical point of view since it does not require the inversion of an operator. Consider the equation  $g = Af$  for some  $g \in K$ . It can be rewritten as follows:

$$f = f + A^*(g - Af). \quad (3.12)$$

The Landweber iterative method is constructed in the following way. Define  $\hat{f}_0 = 0$ . Then, for all  $t \in \mathbb{N}$ ,  $t \geq 1$ , set:

$$\hat{f}_t = \hat{f}_{t-1} + A^*(g - A\hat{f}_{t-1}). \quad (3.13)$$

The number  $t$  of iterations plays the role of the regularization parameter. With the model (1.1), replace  $g$  by  $Y$ . In the SVD setting, the Landweber iterative estimator is:

$$\hat{f}_t = \sum_{k=1}^{+\infty} \lambda_k(t) b_k^{-1} y_k \phi_k, \text{ with } \lambda_k(t) = (1 - (1 - b_k^2)^t), \forall k \in \mathbb{N}.$$

The estimator  $\hat{f}_t$  is defined only if  $\|A\| \leq 1$  (i.e.  $\max_k b_k \leq 1$ ). Otherwise, the method can easily be modified via the introduction of a relaxation parameter in (3.12). The qualification of the Landweber method is  $\infty$ , i.e. there is no restriction on the regularity  $\mu$ .

Using simple algebra, it is possible to prove that  $\sigma^2(t) \simeq \epsilon^4(n_t)^{4\beta+1}$  with  $n_t = t^{1/2\beta}$  and:

$$\begin{aligned} \max_k \lambda_k^2(t) b_k^{-2} &= \max_{x \in (0,1)} \frac{(1 - (1 - x)^t)^2}{x}, \\ &\leq \max_{x \in (0,1)} s'(x), \text{ with } s(x) = 1 - (1 - x)^t, \\ &= t \end{aligned}$$

Hence, (3.1) holds with  $d = 1/2\beta$ . The proof of (3.2) and (3.3) follows essentially the same lines.

For the sake of convenience, only four regularization schemes satisfying the polynomial hypotheses are presented in this paper. The families of estimators covered by these assumptions is certainly larger.

In order to conclude these examples, a few words on the  $\nu$ -methods and the Pinsker estimators. These filters are slightly different from the previous one since both the regularization parameter and the structure of the estimator may depend on the regularity of the function  $f$ . For instance, a Pinsker filter is of the form  $\lambda_k^P = (1 - c_\epsilon a_k)_+$  for all  $k \in \mathbb{N}$ , where both  $c_\epsilon$  and  $(a_k)_k$  depend on the regularity of  $f$ . It is easy to prove that assumptions (3.1)-(3.3) hold for a Pinsker type sequence of the form  $\lambda_k(t) = (1 - t^{-1}k^a)_+$ , for some fixed  $a > 0$ . It is not so clear for  $(\lambda_k^P)_{k \in \mathbb{N}}$ . The same observation holds for the  $\nu$ -methods. Nevertheless, it seems that the behavior of these estimators is close to the examples presented in this section. Hence, the polynomial hypotheses may certainly be modified in order to take into account these specificities.

## 4 Oracle efficiency

Due to the ill-posedness of the problem, the standard URE method is rather unstable. Therefore, estimate and minimize the hull  $V_\alpha(\theta, \lambda(t))$  (introduced in (3.8)), instead of  $R(\theta, \lambda(t))$ . Remark that:

$$\arg \min_{\lambda \in \Lambda} V_\alpha(\theta, \lambda) = \arg \min_{t \in T} \left\{ \sum_{k=1}^{+\infty} [\lambda_k^2(t) - 2\lambda_k(t)]^2 \theta_k^2 + \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2(t) b_k^{-2} + (1 + \alpha) U_0(\lambda(t)) \right\}. \quad (4.1)$$

For all  $\alpha > 0$ , the right-hand side of (4.1) can be estimated by

$$V_\alpha(y, \lambda(t)) = \sum_{k=1}^{+\infty} [\lambda_k^2(t) - 2\lambda_k(t)] (b_k^{-2} y_k^2 - \epsilon^2 b_k^{-2}) + \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2(t) b_k^{-2} + (1 + \alpha) U_0(\lambda(t)). \quad (4.2)$$

Hence consider:

$$\lambda^* = \arg \min_{\lambda \in \Lambda} V_\alpha(y, \lambda), \quad (4.3)$$

and denote by  $\theta^*$  the associated estimator. This approach corresponds in fact to a penalized URE method. The penalty  $(1 + \alpha) U_0(\lambda(t))$  is explicitly computable via Monte-Carlo approximations. The performances of  $\theta^*$  are summarized in the following theorem.

**Theorem 2** *Let  $\theta^*$  the estimator defined in (4.3) with  $\alpha > 1$ . Assume that there exists a positive constant  $C_T$  such that, uniformly in  $t \in T$ :*

$$\sum_{k=1}^{+\infty} \lambda_k^2(t) b_k^{-4} \leq C_T \sum_{k=1}^{+\infty} \lambda_k^4(t) b_k^{-4}. \quad (4.4)$$

*Then, there exist  $B_1, D_1$  and  $\gamma_1 > 0$  independent of  $\epsilon$ , such that, for all  $\lambda \in \Lambda$ , (i.e.  $t \in T$ ) and  $0 < \gamma < \gamma_1$ ,*

$$\mathbb{E}_\theta \|\theta^* - \theta\|^2 \leq (1 + B_1 \gamma) R_\alpha(\theta, \lambda(t)) + D_1 \epsilon^2 \gamma^{-1} L_T \omega(\gamma^{-2} L_T) + \frac{D_1 \epsilon^2}{(\alpha - 1)},$$

*where  $L_T$  is defined in (3.7) and*

$$R_\alpha(\theta, \lambda(t)) = \sum_{k=1}^{+\infty} (1 - \lambda_k(t))^2 \theta_k^2 + \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2(t) b_k^{-2} + (1 + \alpha) U_0(\lambda(t)). \quad (4.5)$$

The proof is presented in Section 5. It is based on Theorem 1 and on the theory of ordered processes (summarized in Section 6). Clearly, inequality (4.4) is verified for projection estimators. This is also the case for Tikhonov estimators, provided  $a > 0$ , Tikhonov iterated procedure ( $j > 1$ ) and for the Landweber iterative method. The constant  $C_T$  has a real impact on the quality of estimation. Indeed, the penalty controls the stochastic terms in the loss and in  $V_\alpha(y, \lambda)$ . For all  $\lambda \in \Lambda$ , the variance of these terms is of order  $\epsilon^4 \sum_k \lambda_k^4 b_k^{-4}$  and  $\epsilon^4 \sum_k \lambda_k^2 b_k^{-4}$ , respectively. If  $C_T$  is too large, the penalty is not sufficient for  $V_\alpha(y, \lambda)$ . The efficiency of  $\theta^*$  will not necessarily be improved compared to the standard URE method.

In Theorem 2, we require  $\alpha$  to be greater than 1. It seems that this condition is too restrictive. Indeed, acceptable bounds for the risk are available for  $\alpha < 1$ . This needs more precision in the proofs. The obtained oracle inequality is not sharp: one cannot consider  $\gamma \rightarrow 0$  as  $\epsilon \rightarrow 0$  in this case. On the other hand, large choices are related to a poor efficiency since  $R_\alpha(\theta, \lambda)$  will be significantly greater than the quadratic risk  $R(\theta, \lambda)$ . From the proofs, it is possible to see that  $\alpha$  has a small influence on the quality of estimation. In such a situation, a choice of  $\alpha$  close to 2 seems to be reasonable. We refer to [9] for a complete discussion concerning this choice of  $\alpha$ .

Several bounds in the proof of Theorem 2 are derived from the theory of ordered processes. These results can easily be applied to the unbiased risk estimator, called  $\tilde{\theta}$  in the following.

**Theorem 3** *Assume that (3.1)-(3.3) and inequality (4.4) hold. Then, there exists  $B_2, D_2$  and  $\gamma_2 > 0$  independent of  $\epsilon$ , such that, for all  $\lambda \in \Lambda$ , (i.e.  $t \in T$ ) and  $0 < \gamma < \gamma_2$ ,*

$$\mathbb{E}_\theta \|\tilde{\theta} - \theta\|^2 \leq (1 + B_2\gamma)R(\theta, \lambda(t)) + D_2\epsilon^2\gamma^{-1}L_T\omega(\gamma^{-2}L_T),$$

where  $R(\theta, \lambda(t))$  and  $L_T$  are defined in (1.5) and (3.7) respectively.

In the oracle inequality of [8], the term  $L_T$  is of order  $\log N$  where  $N$  denotes the size of the family  $\Lambda$ , i.e. the number of considered parameters. Thanks to the theory of ordered processes, the oracle inequalities of Theorems 4.1 and 4.2 are free of assumptions on the size of  $\Lambda$ . In particular, there is no theoretical price to pay for considering large collection of estimators. Moreover, the obtained results concerns continuous intervals for the regularization parameter of the form  $T = [t_{min}; t_{max}]$ .

We have proved that the RHM method initiated by [9] can be enhanced by using more precise estimators like Tikhonov or Landweber iterative methods. The principle of risk hull minimization is not restricted to projection procedures. This generalization is mainly due to recent developments on ordered processes. Section 6 presents the main lines of this theory. It contains important results from [20] and [7] and some additional lemmas useful for the proofs of Theorems 1-3.

Comparing different oracle inequalities in a transparent way is always a very difficult task. The constants and residual terms are derived from successive upper bounds. Nevertheless, the URE method does not take into account the variability of the problem. Hence, the RHM procedure may lead to better performances. The differences between the URE and RHM procedures will be certainly significant for the cases where the variability of the sequence space model is the most important, e.g. when the signal to noise ratio is small or the degree of ill-posedness increases.

In order to conclude this paper, some words on the numerical implementation of the RHM method. The term  $U_0(\lambda(t))$  may be approximated by using Monte-Carlo replications. In this case, the penalty will not be exactly the same than in Theorems 1 and 2. This approximation is equivalent to a perturbation of  $\alpha$ . From proofs of Section 5, it is clear that a small variation in  $\alpha$  does not affect the efficiency of  $\theta^*$ . Therefore, the numerical approximation of  $(1 + \alpha)U_0(\lambda(t))$  is pertinent from a theoretical point of view. Remark also that the lower bound obtained in Lemma 3.1 and used in the most part of the proofs may be used for a numerical implementation.

From a practical point of view, given a regularization scheme and an interval  $T$  for the parameter, finding  $\lambda^*$  may be a difficult task since (4.3) is a non-convex minimization problem. Remark that this inconvenience occurs also for the URE method. A possible outcome is then to consider a finite grid  $\tilde{T}$  and to construct  $\lambda^*$  on  $\tilde{T}$ . From Theorems 4.1 and 4.2, no restriction occurs on the size of the grid. Hence, the limitations are only due to practical considerations (e.g. computer performances). In order to deal with continuous intervals, an interesting alternative would be to develop and use a simulated annealing type algorithm for this setting.

## 5 Proofs

### 5.1 Proof of Lemma 1

The proof follows the same lines of [9]. We shall sometimes omit some technical steps since they have already been developed.

Let  $\eta(t)$ ,  $t \geq 0$ , the stochastic process defined in (3.4). For all  $t \in T$ , introduce:

$$K(t) = \frac{\eta(t)}{\sqrt{2\sigma^2(t)}} \text{ and } u_1(t) = \sqrt{\log\left(\frac{\sigma^2(t)}{2\pi\epsilon^4}\right)}. \quad (5.1)$$

For all  $t \geq 0$ , the function  $x \mapsto \mathbb{E}_\theta K(t) \mathbf{1}_{\{K(t) \geq x\}}$  is monotone decreasing. In order to prove Lemma 1, it suffices to show that:

$$\mathbb{E}_\theta K(t) \mathbf{1}_{\{K(t) \geq u_1(t)\}} \geq \frac{\epsilon^2}{\sqrt{2\sigma^2(t)}}, \quad (5.2)$$

at least for sufficiently large  $t$ . Using integration by part,

$$\mathbb{E}_\theta K(t) \mathbf{1}_{\{K(t) \geq u_1(t)\}} \geq u_1(t) P(K(t) \geq u_1(t)) + \int_{u_1(t)}^{u_1(t)+1} P(K(t) \geq x) dx. \quad (5.3)$$

First, study the characteristic function  $\Upsilon$  of the process  $K(t)$ ,  $t \geq 0$ . For all  $s \in \mathbb{R}$ , using (3.2):

$$|\Upsilon_t(s)| = |\mathbb{E} e^{isK(t)}| \leq \exp \left[ -\frac{1}{4} \sum_{l=1}^{+\infty} \log \left( 1 + \frac{2s^2 \epsilon^4 \lambda_l^4(t) b_l^{-4}}{\sigma^2(t)} \right) \right] \leq \left( 1 + \frac{C_1 s^2}{t^d} \right)^{-t^d}.$$

Setting  $x = \sqrt{t^d/C_1}$ , we obtain:

$$\int_{|s| \geq x} |\Upsilon_s(t)| ds = \int_{x \leq |s| \leq \sqrt{2C_1/t^d}} |\Upsilon_s(t)| ds + \int_{\sqrt{2C_1/t^d} \leq |s|} |\Upsilon_s(t)| ds \leq \exp(-Ct^d). \quad (5.4)$$

Now consider the case  $|s| \leq \sqrt{t^d/C_1}$ . Using a Taylor expansion:

$$\Upsilon_t(s) = \exp \left[ -\frac{s^2}{2} + \sum_{l=3}^{M-1} \frac{(-i)^l 2^{l/2} R_l(t) s^l}{l} + O \left( \frac{CR_M(t) 2^{M/2} s^M}{M} \right) \right],$$

where  $M \in \mathbb{N}$  and for all  $l \in \mathbb{N}$ :

$$R_l(t) = (\sigma^2(t))^{-l/2} \epsilon^{2l} \sum_{p=1}^{+\infty} \lambda_p^{2l}(t) b_p^{-2l}.$$

The behavior of  $R_l(t)$ ,  $l \in \mathbb{N}$  is controlled by (3.3). Expanding  $\Upsilon_t(s) \exp(s^2/2)$  into a Taylor series, construct the following approximation of  $\Upsilon_t(s)$ :

$$\Upsilon_t^M(s) = \exp \left( -\frac{t^2}{2} \right) \left[ 1 + t^d \sum_{l=3}^{M-1} Q_M(l, t) \left( \frac{is}{\sqrt{t^d}} \right)^l \right],$$

where  $Q_M(l, t)$ ,  $l = 3 \dots M$  are function uniformly bounded in  $l$  and  $t$ . Therefore, the probability  $P(K(t) > x)$  can be approximated by:

$$P_t^M(x) = \phi(x) - \frac{k}{\sqrt{2\pi}} \sum_{s=3}^{M-1} (-1)^s Q_M(s, k) k^{-s/2} \frac{d^{s-1}}{dx^{s-1}} \exp \left( -\frac{x^2}{2} \right), \quad (5.5)$$

where  $\phi(x)$  denotes the repartition function of a standard Gaussian random variable at the point  $x$ . Using Parseval identity, (5.4) and (5.5), we prove that:

$$|P(K(t) > x) - P_t^M(x)| \leq \frac{C}{(t^d)^{M/2}}, \quad (5.6)$$

for some positive constant  $C$ : see [9] for more details. Then, using (5.1)-(5.6), integration by part and choosing  $M$  large enough:

$$\begin{aligned}
\mathbb{E}K(t)\mathbf{1}_{\{K(t) \geq u_1(t)\}} &\geq u_1(t)\phi(u_1(t)) + \int_{u_1(t)}^{+\infty} \phi(x)dx \\
&\quad - \int_{u_1(t)+1}^{+\infty} \phi(x)dx - (1+u_1(t)) \max_{u_1(t) \leq x \leq u_1(t)+1} |P_t^M(x) - \phi(x)| - \frac{Cu_1(t)}{(t^d)^{M/2}}, \\
&\geq \frac{1}{\sqrt{2\pi}} e^{-u_1(t)^2/2} - C e^{-(u_1(t)+1)^2/2} - (1+u_1(t)) \max_{u_1(t) \leq x \leq u_1(t)+1} |P_t^M(x) - \phi(x)| - \frac{Cu_1(t)}{(t^d)^{M/2}}, \\
&\geq \frac{\epsilon^2}{\sqrt{\sigma^2(t)}} + o\left(\frac{\epsilon^2}{\sqrt{2\sigma^2(t)}}\right), \text{ as } t \rightarrow +\infty.
\end{aligned}$$

Indeed, (3.1) and (5.1) provide  $u_1(t) \simeq C\sqrt{\log(t)}$  for some positive constant  $C > 0$ . Hence, (5.1) is satisfied. This concludes the proof of Lemma 1.  $\square$

## 5.2 Proof of Theorem 3.2

Remark that:

$$\begin{aligned}
&\mathbb{E}_\theta \sup_{\lambda \in \Lambda} \left\{ \|\hat{\theta}_\lambda - \theta\|^2 - V_\alpha(\theta, \lambda) \right\} \\
&= \mathbb{E}_\theta \sup_{\lambda \in \Lambda} \left[ \sum_{k=1}^{+\infty} (\lambda_k b_k^{-1} y_k - \theta_k)^2 - V_\alpha(\theta, \lambda) \right], \\
&= \mathbb{E}_\theta \left[ \sum_{k=1}^{+\infty} (1 - \bar{\lambda}_k)^2 \theta_k^2 + \epsilon^2 \sum_{k=1}^{+\infty} \bar{\lambda}_k^2 b_k^{-2} \xi_k^2 + 2 \sum_{k=1}^{+\infty} (\bar{\lambda}_k - 1) \bar{\lambda}_k \theta_k \epsilon b_k^{-1} \xi_k - V_\alpha(\theta, \lambda) \right],
\end{aligned}$$

with,

$$\bar{\lambda} = \lambda(\bar{t}) = \arg \sup_{\lambda \in \Lambda} \left\{ \|\hat{\theta}_\lambda - \theta\|^2 - V_\alpha(\theta, \lambda) \right\}.$$

Use the following decomposition:

$$\begin{aligned}
2\mathbb{E}_\theta \sum_{k=1}^{+\infty} (\bar{\lambda}_k - 1) \bar{\lambda}_k \theta_k \epsilon b_k^{-1} \xi_k &= 2\mathbb{E}_\theta \sum_{k=1}^{+\infty} \{\bar{\lambda}_k^2 - \bar{\lambda}_k\} \theta_k \epsilon b_k^{-1} \xi_k, \\
&= \mathbb{E}_\theta \sum_{k=1}^{+\infty} \bar{\lambda}_k^2 \theta_k \epsilon b_k^{-1} \xi_k + \mathbb{E}_\theta \sum_{k=1}^{+\infty} \{\bar{\lambda}_k^2 - 2\bar{\lambda}_k\} \theta_k \epsilon b_k^{-1} \xi_k, \\
&= \mathbb{E}_\theta \sum_{k=1}^{+\infty} \bar{\lambda}_k^2 \theta_k \epsilon b_k^{-1} \xi_k + \mathbb{E}_\theta \sum_{k=1}^{+\infty} (1 - \bar{\lambda}_k)^2 \theta_k \epsilon b_k^{-1} \xi_k.
\end{aligned}$$

Corollary 1 in Section 6 provides, for all  $\gamma > 0$ :

$$2\mathbb{E}_\theta \sum_{k=1}^{+\infty} (\bar{\lambda}_k - 1) \bar{\lambda}_k \theta_k \epsilon b_k^{-1} \xi_k \leq \gamma \mathbb{E}_\theta R(\theta, \bar{\lambda}) + \gamma R(\theta, \lambda^0) + C\epsilon^2 \gamma^{-1} L_T \omega(\gamma^{-2} L_T),$$

for some positive constant  $C > 0$ . Therefore,

$$\begin{aligned} & \mathbb{E}_\theta \sup_{\lambda \in \Lambda} \{ \|\hat{\theta}_\lambda - \theta\|^2 - V_\alpha(\theta, \lambda) \} \\ & \leq \mathbb{E}_\theta \sup_{\lambda \in \Lambda} \left[ (1 + \gamma) \sum_{k=1}^{+\infty} (1 - \lambda_k^2) \theta_k^2 + \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2 b_k^{-2} \xi_k^2 + \gamma \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2 b_k^{-2} \right. \\ & \quad \left. + C \gamma^{-1} \epsilon^2 L_T \omega(\gamma^{-2} L_T) + \gamma R(\theta, \lambda^0) - V_\alpha(\theta, \lambda) \right], \end{aligned} \quad (5.7)$$

$$\begin{aligned} & \leq \mathbb{E} \sup_{\lambda \in \Lambda} \left[ \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2 b_k^{-2} (\xi_k^2 - 1) - (1 + \gamma)(1 + \alpha) U_0(\lambda) - \frac{C^* \epsilon^2}{\alpha} \right], \\ & \leq \mathbb{E} \sup_{\lambda \in \Lambda} \left[ \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2 b_k^{-2} (\xi_k^2 - 1) - (1 + \alpha) U_0(\lambda) \right] - \frac{C^* \epsilon^2}{\alpha}. \end{aligned} \quad (5.8)$$

Let  $\sigma^2(t)$  and  $\eta(t), t > 0$  the quantities defined in (3.1) and (3.4), respectively. Recall that the function  $t \mapsto \sigma^2(t)$  is monotone non-decreasing since it denotes the variance of the ordered process  $\eta(t)$ ,  $t \geq 0$  up to some constant. Without loss of generality, we consider that  $T = ]0; +\infty[$  and  $\sigma^2(t) \rightarrow +\infty$  as  $t \rightarrow +\infty$ . Let  $(t_s)_{s \in \mathbb{N}}$  be a grid in  $\mathbb{R}$  which will be chosen later. Using Lemma 1, it is easy to see that:

$$\begin{aligned} & \mathbb{E} \sup_{\lambda \in \Lambda} \left[ \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2 b_k^{-2} (\xi_k^2 - 1) - (1 + \alpha) U_0(\lambda) \right] \\ & \leq \sum_{s=0}^{+\infty} \mathbb{E} \sup_{t \in [t_s, t_{s+1}]} \left[ \eta(t) - (1 + \alpha) \sqrt{2\sigma^2(t) \log(C\epsilon^{-4}\sigma^2(t))} \right]_+, \end{aligned}$$

for some positive constant  $C$ . For all  $x \in \mathbb{R}$ , the notation  $x_+$  denotes the positive part of  $x$ , i.e.  $x_+ = x \mathbf{1}_{\{x > 0\}}$ . Let  $s \in \mathbb{N}$  be fixed:

$$\begin{aligned} & \mathbb{E} \sup_{t \in [t_s, t_{s+1}]} \left[ \eta(t) - (1 + \alpha) \sqrt{2\sigma^2(t) \log(C\epsilon^{-4}\sigma^2(t))} \right]_+ \\ & = \mathbb{E} \sup_{t \in [t_s, t_{s+1}]} \left[ \eta(t_s) - \eta(t_s) + \eta(t) - (1 + \alpha) \sqrt{2\sigma^2(t) \log(C\epsilon^{-4}\sigma^2(t))} \right]_+ \\ & \leq \mathbb{E} \left[ \eta(t_s) - (1 + \alpha) \sqrt{2\sigma^2(t_s) \log(C\epsilon^{-4}\sigma^2(t_s))} + \mathcal{E}(t_s) \right]_+, \end{aligned} \quad (5.9)$$

where, for all  $s \in \mathbb{N}$ ,

$$\mathcal{E}(t_s) = \sup_{t \in [t_s, t_{s+1}]} \{ \eta(t) - \eta(t_s) \}. \quad (5.10)$$

Since the variables  $\eta(t_s)$  and  $\mathcal{E}(t_s)$  are not independent, the same method as [9] cannot be applied. Instead, remark that for all  $0 < \mu < 1$ , using (5.9),

$$\begin{aligned} & \mathbb{E} \sup_{t \in [t_s, t_{s+1}]} \left[ \eta(t) - (1 + \alpha) \sqrt{2\sigma^2(t) \log(C\epsilon^{-4}\sigma^2(t))} \right]_+ \\ & \leq \int_{B_s}^{+\infty} P(\eta(t_s) + \mathcal{E}(t_s) \geq x) dx, \\ & \leq \int_{B_s}^{+\infty} P(\eta(t_s) \geq \mu x) dx + \int_{B_s}^{+\infty} P(\mathcal{E}(t_s) \geq (1 - \mu)x) dx = A_1 + A_2, \end{aligned} \quad (5.11)$$

with,

$$B_s = (1 + \alpha) \sqrt{2\sigma^2(t_s) \log(C\epsilon^{-4}\sigma^2(t_s))}, \quad \forall s \in \mathbb{N}. \quad (5.12)$$

We are first interested in the study of  $A_2$ . For all  $p > 1$ , using a Markov inequality,

$$\begin{aligned}
A_2 &= \int_{B_s}^{+\infty} P(\mathcal{E}(t_s) \geq (1-\mu)x) dx = \int_{B_s}^{+\infty} P\left(\sup_{t \in [t_s, t_{s+1}]} \{\eta(t) - \eta(t_s)\} \geq (1-\mu)x\right) dx, \\
&\leq \frac{1}{(1-\mu)^p} \int_{B_s}^{+\infty} \frac{\mathbb{E} \sup_{t \in [t_s, t_{s+1}]} |\eta(t) - \eta(t_s)|^p}{x^p} dx, \\
&= \frac{1}{(1-\mu)^p} \mathbb{E} \sup_{t \in [t_s, t_{s+1}]} |\eta(t) - \eta(t_s)|^p \frac{1}{(p-1)B_s^{p-1}}.
\end{aligned} \tag{5.13}$$

The process  $\eta(t)$ ,  $t > 0$  is ordered (see Section 6). By Lemma 2, there exists a positive constant  $C > 0$  independent of  $\epsilon$  and  $t$ , such that, for all  $p > 0$ :

$$\mathbb{E} \sup_{t \in [t_s, t_{s+1}]} |\eta(t) - \eta(t_s)|^p \leq C^p p^p (\sigma^2(t_{s+1}) - \sigma^2(t_s))^{p/2}. \tag{5.14}$$

Therefore, using (5.12)-(5.14), we obtain:

$$A_2 \leq \left(\frac{C}{(1-\mu)(1+\alpha)}\right)^p \frac{p^p}{p-1} \frac{(\sigma^2(t_{s+1}) - \sigma^2(t_s))^{p/2}}{[\sigma^2(t_s) \log(C\epsilon^{-4}\sigma^2(t_s))]^{(p-1)/2}}. \tag{5.15}$$

Now, choose the grid  $(t_s)_{s \in \mathbb{N}}$  as follows:

$$\sigma^2(t_s) = \epsilon^4 \left(1 + \frac{1}{s^\nu}\right)^s, \quad \forall s \in \mathbb{N}, \tag{5.16}$$

where  $\nu \in ]3/4, 1[$ . This choice is a trade off between an exponential grid similar to [9] and polynomial grids of [7]. Remark that:

$$\sigma^2(t_{s+1}) - \sigma^2(t_s) = \epsilon^4 \left(1 + \frac{1}{(s+1)^\nu}\right)^{s+1} - \epsilon^4 \left(1 + \frac{1}{s^\nu}\right)^s \leq \epsilon^4 \left(1 + \frac{1}{s^\nu}\right)^s \frac{1}{s^\nu}. \tag{5.17}$$

Moreover, for all  $s \in \mathbb{N}$ , using Taylor formula,

$$\log(C\epsilon^{-4}\sigma^2(t_s)) = s \left[ \log(C^{1/s}) + \log\left(1 + \frac{1}{s^\nu}\right) \right] \geq s \log\left(1 + \frac{1}{s^\nu}\right) \geq Cs^{1-\nu}. \tag{5.18}$$

Hence, (5.15)-(5.18) yield:

$$\begin{aligned}
A_2 &\leq \epsilon^2 \left[ \frac{C}{(1-\mu)(1+\alpha)} \right]^p \frac{p^p}{p-1} \frac{(s^{-\nu})^{p/2} (1 + s^{-\nu})^{sp/2}}{(1 + s^{-\nu})^{s(p-1)/2} (s^{1-\nu})^{(p-1)/2}}, \\
&= \epsilon^2 \left[ \frac{C}{(1-\mu)(1+\alpha)} \right]^p \frac{p^p}{p-1} \frac{(1 + s^{-\nu})^{s/2}}{(s^\nu)^{p/2} (s^{1-\nu})^{(p-1)/2}}, \\
&= \epsilon^2 \left[ \frac{C}{(1-\mu)(1+\alpha)} \right]^p \frac{p^p}{p-1} s^{(1-\nu)/2} \times \frac{(1 + s^{-\nu})^{s/2}}{s^{p/2}}.
\end{aligned}$$

For all  $s \in \mathbb{N}$ , with simple algebra,

$$(1 + s^{-\nu})^{s^\nu} = \exp\{s^\nu \log(1 + s^{-\nu})\} \simeq \exp(s^\nu \times s^{-\nu}) = O(1), \text{ as } s \rightarrow +\infty.$$

Therefore:

$$(1 + s^{-\nu})^{s/2} = \exp\left\{\frac{s^{1-\nu}}{2} \log\left((1 + s^{-\nu})^{s^\nu}\right)\right\} \leq Cs^{1-\nu},$$

for some positive constant  $C > 0$ . We eventually obtain:

$$A_2 \leq \epsilon^2 g(\mu, \alpha)^p \frac{s^{(1-\nu)/2}}{p-1} \times \frac{p^p}{\sqrt{s^p}} Cs^{1-\nu}, \text{ with } g(\mu, \alpha) = \left[ \frac{C}{(1-\mu)(1+\alpha)} \right]. \tag{5.19}$$

Now, choose the parameter  $p$  in a properly way. For example, set  $p = s^{1/4}$ . Then:

$$\begin{aligned}
A_2 &\leq \epsilon^2 s^{\frac{1-\nu}{2}-\frac{1}{4}} g(\mu, \alpha)^{s^{1/4}} \left( \frac{s^{1/4}}{s^{1/2}} \right)^{s^{1/4}} C^{s^{1-\nu}}, \\
&= \epsilon^2 s^{\frac{1}{4}-\frac{\nu}{2}} g(\mu, \alpha)^{s^{1/4}} (s^{1/4})^{-s^{1/4}} C^{s^{1-\nu}}, \\
&= \epsilon^2 s^{\frac{1}{4}-\frac{\nu}{2}} \exp \left\{ s^{1/4} \log(g(\mu, \alpha)) + s^{1-\nu} \log C - s^{1/4} \log(s^{1/4}) \right\}, \\
&= \epsilon^2 s^{\frac{1}{4}-\frac{\nu}{2}} \exp \left\{ s^{1/4} \left[ \log(g(\mu, \alpha)) + s^{1-\nu-1/4} \log C - \frac{1}{4} \log(s) \right] \right\}.
\end{aligned}$$

It is then easy to see that, provided  $\nu \in ]3/4, 1[$ ,

$$\log(g(\mu, \alpha)) + s^{1-\nu-1/4} \log C - \frac{1}{4} \log(s) < 0,$$

at least for  $s$  large enough. Therefore, for all  $\mu \in (0, 1)$ :

$$A_2 \leq \epsilon^2 C e^{-s^{1/4}} \text{ and } \sum_{s=1}^{+\infty} \int_{B_s}^{+\infty} P(\mathcal{E}(t_s) \geq (1-\mu)x) dx \leq C_1^* \epsilon^2, \quad (5.20)$$

where  $C_1^*$  is a positive constant independent of  $\epsilon$ .

We are now interested in the study of  $A_1$  in inequality (5.11). For all  $\delta > 0$ , using a Markov inequality:

$$\begin{aligned}
A_1 &= \int_{B_s}^{+\infty} P(\eta(t_s) \geq \mu x) dx, \\
&= \int_{B_s}^{+\infty} P(\exp(\delta \eta(t_s)) \geq \exp(\delta \mu x)) dx, \\
&\leq \int_{B_s}^{+\infty} \frac{\mathbb{E} \exp\{\delta \eta(t_s)\}}{\exp\{\delta \mu x\}} dx = \frac{1}{\delta \mu} \mathbb{E} \exp\{\delta \eta(t_s)\} e^{-\delta \mu B_s}.
\end{aligned} \quad (5.21)$$

Inequality (4.14) of [9] yields:

$$\mathbb{E} \exp\{\delta \eta(t_s)\} \leq \exp \left\{ \delta^2 \sigma^2(t_s) + 4\delta^3 \sum_{k=1}^{+\infty} \frac{\epsilon^6 \lambda_k(t_s)^6 b_k^{-6}}{(1 - 2\delta \lambda_k^2(t_s) b_k^{-2})_+} \right\}. \quad (5.22)$$

Setting,

$$\delta = \sqrt{\frac{\log(C\epsilon^{-4}\sigma^2(t_s))}{2\sigma^2(t_s)}}, \quad (5.23)$$

we obtain,

$$\begin{aligned}
A_1 &\leq C\mu^{-1} \sqrt{\frac{2\sigma^2(t_s)}{\log(C\epsilon^{-4}\sigma^2(t_s))}} \exp \left\{ \frac{1}{2} \log(C\epsilon^{-4}\sigma^2(t_s)) \right\} \times e^{-\mu(1+\alpha) \log(C\epsilon^{-4}\sigma^2(t_s))}, \\
&\leq C\mu^{-1} \epsilon^2 \sqrt{\frac{1}{\log(C\epsilon^{-4}\sigma^2(t_s))}} \exp \left\{ -[\mu(1+\alpha) - 1] \log(C\epsilon^{-4}\sigma^2(t_s)) \right\}.
\end{aligned}$$

Indeed, using (3.1)-(3.3) and (5.23):

$$\delta^3 \sum_{k=1}^{+\infty} \frac{\lambda_k^6(t_s) b_k^{-6}}{(1 - 2\delta \lambda_k^2(t_s) b_k^{-2})_+} \rightarrow 0, \text{ as } s \rightarrow +\infty.$$

With (5.16) and (5.18), we eventually obtain:

$$\begin{aligned}
\sum_{s=1}^{+\infty} \int_{B_s}^{+\infty} P(\eta(t_s) \geq \mu x) dx &\leq \sum_{s=1}^{+\infty} \frac{C\epsilon^2}{s^{(1-\nu)/2}} \left[ e^{-s \log(1+s^{-\nu})} \right]^{\mu(1+\alpha)-1}, \\
&\leq \sum_{s=1}^{+\infty} \frac{C\epsilon^2}{s^{(1-\nu)/2}} \exp\{-C(\mu(1+\alpha) - 1)s^{1-\nu}\} \leq \frac{C_2^* \epsilon^2}{\alpha}, \quad (5.24)
\end{aligned}$$



setting for example  $\mu = (1 + \alpha/2)/(1 + \alpha)$ . Therefore,

$$\mathbb{E} \sup_{\lambda \in \Lambda} \left[ \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2 b_k^{-2} (\xi_k^2 - 1) - (1 + \alpha) U_0(\lambda) \right] \leq \frac{C^* \epsilon^2}{\alpha}. \quad (5.25)$$

Theorem 1 is obtained using (5.8) and (5.25).  $\square$

### 5.3 Proof of Theorem 2

Let  $\mu > 0$ . By Theorem 1,  $V_\mu(\theta, \lambda)$  is a risk hull. Therefore,

$$\mathbb{E}_\theta \|\theta^* - \theta\|^2 \leq \mathbb{E}_\theta V_\mu(\theta, \lambda^*). \quad (5.26)$$

Moreover, (4.3) provides that,  $\forall \lambda \in \Lambda$ ,

$$\mathbb{E}_\theta V_\alpha(y, \lambda^*) \leq \mathbb{E}_\theta V_\alpha(y, \lambda). \quad (5.27)$$

We will combine inequalities (5.26) and (5.27). First, rewrite  $V_\alpha(y, \lambda^*)$  in terms of  $V_\mu(\theta, \lambda^*)$ . By simple algebra:

$$\begin{aligned} V_\alpha(y, \lambda^*) &= \sum_{k=1}^{+\infty} \{(\lambda_k^*)^2 - 2\lambda_k^*\} (b_k^{-2} y_k^2 - \epsilon^2 b_k^{-2}) + \epsilon^2 \sum_{k=1}^{+\infty} (\lambda_k^*)^2 b_k^{-2} + (1 + \alpha) U_0(\lambda^*), \\ &= \sum_{k=1}^{+\infty} \{(1 - \lambda_k^*)^2 \theta_k^2 + \epsilon^2 (\lambda_k^*)^2 b_k^{-2}\} + (1 + \mu) U_0(\lambda^*) + (\alpha - \mu) U_0(\lambda^*), \\ &\quad + 2 \sum_{k=1}^{+\infty} \{(\lambda_k^*)^2 - 2\lambda_k^*\} \theta_k \epsilon b_k^{-1} \xi_k + \sum_{k=1}^{+\infty} \{(\lambda_k^*)^2 - 2\lambda_k^*\} \epsilon^2 b_k^{-2} (\xi_k^2 - 1) - \|\theta\|^2, \\ &= R_\mu(\theta, \lambda^*) + (\alpha - \mu) U_0(\lambda^*) - \|\theta\|^2 \\ &\quad + 2 \sum_{k=1}^{+\infty} \{(\lambda_k^*)^2 - 2\lambda_k^*\} \theta_k \epsilon b_k^{-1} \xi_k + \sum_{k=1}^{+\infty} \{(\lambda_k^*)^2 - 2\lambda_k^*\} \epsilon^2 b_k^{-2} (\xi_k^2 - 1), \end{aligned} \quad (5.28)$$

where  $R_\mu(\theta, \lambda^*)$  is defined in (4.5). Using (5.27) and (5.28), for all  $\lambda \in \Lambda$ ,

$$\begin{aligned} \mathbb{E}_\theta R_\mu(\theta, \lambda^*) &\leq \mathbb{E}_\theta V_\alpha(y, \lambda) - (\alpha - \mu) \mathbb{E}_\theta U_0(\lambda^*) + \|\theta\|^2 \\ &\quad - 2 \mathbb{E}_\theta \sum_{k=1}^{+\infty} \{(\lambda_k^*)^2 - 2\lambda_k^*\} \theta_k \epsilon b_k^{-1} \xi_k - \mathbb{E}_\theta \sum_{k=1}^{+\infty} \{(\lambda_k^*)^2 - 2\lambda_k^*\} \epsilon^2 b_k^{-2} (\xi_k^2 - 1). \end{aligned} \quad (5.29)$$

It is then easy to see that, for all  $\lambda \in \Lambda$ ,

$$\mathbb{E}_\theta V_\alpha(y, \lambda) = R_\alpha(\theta, \lambda) - \|\theta\|^2.$$

This yields:

$$\begin{aligned} \mathbb{E}_\theta R_\mu(\theta, \lambda^*) &\leq R_\alpha(\theta, \lambda) - 2 \mathbb{E}_\theta \sum_{k=1}^{+\infty} \{(\lambda_k^*)^2 - 2\lambda_k^*\} \theta_k \epsilon b_k^{-1} \xi_k \\ &\quad + \mathbb{E}_\theta \left[ \sum_{k=1}^{+\infty} \{2\lambda_k^* - (\lambda_k^*)^2\} \epsilon^2 b_k^{-2} (\xi_k^2 - 1) - (\alpha - \mu) U_0(\lambda^*) \right]. \end{aligned} \quad (5.30)$$

In order to complete the proof, we bound the two last terms of (5.30). First use Corollary 1 in Section 6. For all  $\gamma > 0$  and for all  $\lambda \in \Lambda$ :

$$\begin{aligned} \mathbb{E}_\theta \sum_{k=1}^{+\infty} \{(\lambda_k^*)^2 - 2\lambda_k^*\} \theta_k \epsilon b_k^{-1} \xi_k &= \mathbb{E}_\theta \sum_{k=1}^{+\infty} (1 - \lambda_k^*)^2 \theta_k \epsilon b_k^{-1} \xi_k \\ &\geq -\gamma \mathbb{E}_\theta R(\theta, \lambda^*) - \gamma R(\theta, \lambda) - C \gamma^{-1} \epsilon^2 L_T \omega(\gamma^{-2} L_T) \end{aligned} \quad (5.31)$$

Therefore, using (5.29)-(5.31), we obtain:

$$\begin{aligned}
& \mathbb{E}_\theta R_\mu(\theta, \lambda^*) \\
& \leq (1 + \gamma)R_\alpha(\theta, \lambda) + \gamma \mathbb{E}_\theta \left\{ \sum_{k=1}^{+\infty} (1 - \lambda_k^*)^2 \theta_k^2 + \epsilon^2 \sum_{k=1}^{+\infty} (\lambda_k^*)^2 b_k^{-2} \right\} \\
& \quad + C\gamma^{-1} \epsilon^2 L_T \omega(\gamma^{-2} L_T) + \mathbb{E}_\theta \left[ \sum_{k=1}^{+\infty} \{2\lambda_k^* - (\lambda_k^*)^2\} \epsilon^2 b_k^{-2} (\xi_k^2 - 1) - (\alpha - \mu) U_0(\lambda^*) \right], \\
& = (1 + \gamma)R_\alpha(\theta, \lambda) + \gamma \mathbb{E}_\theta \left\{ \sum_{k=1}^{+\infty} (1 - \lambda_k^*)^2 \theta_k^2 + \epsilon^2 \sum_{k=1}^{+\infty} (\lambda_k^*)^2 b_k^{-2} \xi_k^2 \right\} + C\gamma^{-1} \epsilon^2 L_T \omega(\gamma^{-2} L_T), \\
& \quad + \mathbb{E}_\theta \left[ \sum_{k=1}^{+\infty} \{2\lambda_k^* - (1 + \gamma)(\lambda_k^*)^2\} \epsilon^2 b_k^{-2} (\xi_k^2 - 1) - (\alpha - \mu) U_0(\lambda^*) \right]. \tag{5.32}
\end{aligned}$$

Now, study the last term in the right-hand side of (5.32). It can be rearranged as follow:

$$\begin{aligned}
& \mathbb{E}_\theta \left[ \sum_{k=1}^{+\infty} \{2\lambda_k^* - (1 + \gamma)(\lambda_k^*)^2\} \epsilon^2 b_k^{-2} (\xi_k^2 - 1) - (\alpha - \mu) U_0(\lambda^*) \right], \\
& = \mathbb{E}_\theta \left[ \epsilon^2 \sum_{k=1}^{+\infty} (1 - \gamma)(\lambda_k^*)^2 b_k^{-2} (\xi_k^2 - 1) - (\alpha - \mu) U_0(\lambda^*) \right] \\
& \quad + 2\mathbb{E}_\theta \sum_{k=1}^{+\infty} \{ \lambda_k^* - (\lambda_k^*)^2 \} \epsilon^2 b_k^{-2} (\xi_k^2 - 1) = T_1 + T_2.
\end{aligned}$$

We are first interested in  $T_1$ . The bound of this term represents the gain on the traditional URE method. Indeed, if the penalty is zero, only less precise bounds are available (see proof of Theorem 3). Remark that if the constant  $C_T$  is large, the term  $T_1$  will be negligible compared to  $T_2$ . The efficiency will not be enhanced by the RHM method. Here, using (5.25):

$$T_1 = (1 - \gamma) \mathbb{E}_\theta \left[ \epsilon^2 \sum_{k=1}^{+\infty} (\lambda_k^*)^2 b_k^{-2} (\xi_k^2 - 1) - \frac{(\alpha - \mu)}{1 - \gamma} U_0(\lambda^*) \right] \leq \frac{(1 - \gamma)^2 C \epsilon^2}{(\alpha - \mu + \gamma - 1)_+}. \tag{5.33}$$

Then study the stochastic term  $T_2$ :

$$T_2 = 2\epsilon^2 \mathbb{E}_\theta \sum_{k=1}^{+\infty} \lambda_k^* b_k^{-2} (\xi_k^2 - 1) - 2\epsilon^2 \mathbb{E}_\theta \sum_{k=1}^{+\infty} (\lambda_k^*)^2 b_k^{-2} (\xi_k^2 - 1). \tag{5.34}$$

For all  $\lambda \in \Lambda$  and  $B > 0$ , using Lemma 3 and the same technics of proof of Corollary 1 (see Section 6) and (6.9):

$$\begin{aligned}
& 2\epsilon^2 \mathbb{E}_\theta \sum_{k=1}^{+\infty} \lambda_k^* b_k^{-2} (\xi_k^2 - 1) \\
& \leq C \log^2 S \times \mathbb{E}_\theta \sqrt{\epsilon^4 \sum_{k=1}^{+\infty} \lambda_k^2(t^*) b_k^{-4}} + \frac{1}{\sqrt{S}} \sqrt{\epsilon^4 \mathbb{E}_\theta \sum_{k=1}^{+\infty} \lambda_k^2(t^*) b_k^{-4}}, \\
& \leq C \log^2 S \times \epsilon^2 \mathbb{E}_\theta \sqrt{\sup_{k \in \mathbb{N}} (\lambda_k^*)^2 b_k^{-2} \sum_{k=1}^{+\infty} \lambda_k^2(t^*) b_k^{-2}} + C \epsilon^2 \sqrt{\mathbb{E}_\theta \sup_{k \in \mathbb{N}} (\lambda_k^*)^2 b_k^{-2} \mathbb{E}_\theta \sum_{k=1}^{+\infty} \lambda_k^2(t^*) b_k^{-2}}, \\
& \leq \frac{\gamma}{2} \epsilon^2 \mathbb{E}_\theta \sum_{k=1}^{+\infty} \lambda_k^2(t^*) b_k^{-2} + C \epsilon^2 \gamma^{-1} \log^2 S \times \mathbb{E}_\theta \sup_{k \in \mathbb{N}} (\lambda_k^*)^2 b_k^{-2}, \\
& \leq \gamma \epsilon^2 \mathbb{E}_\theta \sum_{k=1}^{+\infty} \lambda_k^2(t^*) b_k^{-2} + C \epsilon^2 \gamma^{-1} L_T \omega(\gamma^{-2} L_T). \tag{5.35}
\end{aligned}$$

Indeed, the process

$$\rho(t) = \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k(t) b_k^{-2} (\xi_k^2 - 1), \quad t \geq 0,$$

is ordered. The same bound for the second term in the right hand side of (5.34) occurs since the process  $-\eta(t)$ ,  $t \geq 0$  is also ordered. Using (3.8), (5.7) and (5.32)-(5.35), we eventually find:

$$\begin{aligned} (1 - c\gamma) \mathbb{E}_\theta V_\mu(\theta, \lambda^*) &\leq (1 + \gamma)^2 R_\alpha(\theta, \lambda) + \gamma \mathbb{E}_\theta \left[ \sum_{k=1}^{+\infty} (1 - \lambda_k^*)^2 \theta_k^2 + \epsilon^2 \sum_{k=1}^{+\infty} (\lambda_k^*)^2 b_k^{-2} \xi_k^2 \right] \\ &\quad + C\gamma^{-1} \epsilon^2 L_T \omega(\gamma^{-2} L_T) + \frac{C(1 - \gamma^2) \epsilon^2}{(\alpha - \mu + \gamma - 1)_+} + \frac{C\epsilon^2}{\mu}, \\ &\leq (1 + \gamma)^2 R_\alpha(\theta, \lambda) + C\gamma \mathbb{E}_\theta V_\mu(\theta, \lambda^*) \\ &\quad + C\gamma^{-1} \epsilon^2 L_T \omega(\gamma^{-2} L_T) + \frac{C(1 - \gamma^2) \epsilon^2}{(\alpha - \mu + \gamma - 1)_+} + \frac{C\epsilon^2}{\mu}, \end{aligned} \quad (5.36)$$

for some  $c > 0$ . Combining inequalities (5.26), (5.36), choosing  $\gamma < \gamma_0$  for some  $\gamma_0$ , we eventually obtain:

$$\begin{aligned} \mathbb{E}_\theta \|\theta^* - \theta\|^2 &\leq \mathbb{E}_\theta V_\mu(\theta, \lambda^*), \\ &\leq \left( \frac{(1 + \gamma)^2}{1 - c\gamma} \right) R_\alpha(\theta, \lambda) + \frac{C\epsilon^2}{\mu} + \frac{(1 - \gamma^2) C\epsilon^2}{(\alpha - \mu + \gamma - 1)_+} + C\gamma^{-1} \epsilon^2 L_T \omega(\gamma^{-2} L_T). \end{aligned}$$

Set  $\mu = \gamma$  in order to conclude the proof. □

## 5.4 Proof of Theorem 4.2

The proof essentially follows the same lines of Sections 5.2 and 5.3. Let  $\gamma > 0$  be fixed and  $W(\theta, \lambda)$  defined by:

$$\begin{aligned} W(\theta, \lambda(t)) &= (1 + 2\gamma) \left\{ \sum_{k=1}^{+\infty} (1 - \lambda_k(t))^2 \theta_k^2 + \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2(t) b_k^{-2} \right\} \\ &\quad + \tilde{C} \epsilon^2 \gamma^{-1} L_T \omega(\gamma^{-2} L_T) + \gamma R(\theta, \lambda_0), \end{aligned} \quad (5.37)$$

where  $\tilde{C}$  denotes a positive constant independent of  $\epsilon$ . First remark that using (5.7):

$$\begin{aligned} &\mathbb{E}_\theta \sup_{\lambda \in \Lambda} \left\{ \|\hat{\theta}_\lambda - \theta\|^2 - W(\theta, \lambda) \right\} \\ &\leq \mathbb{E}_\theta \sup_{\lambda \in \Lambda} \left[ (1 + \gamma) \sum_{k=1}^{+\infty} (1 - \lambda_k)^2 \theta_k^2 + \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2 b_k^{-2} \xi_k^2 + \gamma \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2 b_k^{-2} \right. \\ &\quad \left. + C\gamma^{-1} \epsilon^2 \omega(\gamma^{-2} L_T) + \gamma R(\theta, \lambda^0) - W(\theta, \lambda) \right], \\ &= \mathbb{E}_\theta \sup_{t \in T} \left[ (1 + \gamma) \sum_{k=1}^{+\infty} (1 - \lambda_k(t))^2 \theta_k^2 + \eta(t) + (1 + \gamma) \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2(t) b_k^{-2} \right. \\ &\quad \left. + C\gamma^{-1} \epsilon^2 \omega(\gamma^{-2} L_T) + \gamma R(\theta, \lambda^0) - W(\theta, \lambda) \right], \end{aligned}$$

The process  $\eta(t)$ ,  $t \geq 0$  is ordered. Using (5.35):

$$\mathbb{E}_\theta \eta(\tilde{t}) \leq \gamma \epsilon^2 \mathbb{E}_\theta \sum_{k=1}^{+\infty} \lambda_k^2(\tilde{t}) b_k^{-2} + C\epsilon^2 \gamma^{-1} L_T \omega(\gamma^{-2} L_T),$$

for some positive constant  $C$  and  $\gamma > 0$ , with,

$$\lambda(\tilde{t}) = \arg \sup_{t \in T} \left\{ \|\hat{\theta}_\lambda - \theta\|^2 - W(\theta, \lambda) \right\}.$$

This inequality can easily be derived from Lemma 3 and proof of Corollary 1. Hence:

$$\begin{aligned} & \mathbb{E}_\theta \sup_{\lambda \in \Lambda} \left\{ \|\hat{\theta}_\lambda - \theta\|^2 - W(\theta, \lambda) \right\} \\ & \leq \mathbb{E}_\theta \sup_{\lambda \in \Lambda} \left[ (1 + \gamma) \sum_{k=1}^{+\infty} (1 - \lambda_k^2) \theta_k^2 + (1 + 2\gamma) \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k^2 b_k^{-2} \right. \\ & \quad \left. + C\gamma^{-1} \epsilon^2 L_T \omega(\gamma^{-2} L_T) + \gamma R(\theta, \lambda^0) - W(\theta, \lambda) \right] \leq 0. \end{aligned} \quad (5.38)$$

Therefore,  $W(\theta, \lambda)$  can also be considered as a risk hull. Hence, the traditional URE procedure is in some sense a risk hull method. Nevertheless, the hull  $W(\theta, \lambda)$  is rather rough compared to  $V_\alpha(\theta, \lambda)$ . The variability of the problem is neglected.

Using the same principle as in (5.26)-(5.32) and (5.38), we get:

$$\begin{aligned} \mathbb{E}_\theta R(\theta, \lambda^*) & \leq (1 + \gamma) R(\theta, \lambda) + \gamma \mathbb{E}_\theta R(\theta, \lambda^*) + C\gamma^{-1} \epsilon^2 L_T \omega(\gamma^{-2} L_T) \\ & \quad + \mathbb{E}_\theta \sum_{k=1}^{+\infty} \{2\lambda_k^* - (\lambda_k^*)^2\} \epsilon^2 b_k^{-2} (\xi_k^2 - 1), \\ & \leq (1 + \gamma) R(\theta, \lambda) + 2\gamma \mathbb{E}_\theta R(\theta, \lambda^*) + C\gamma^{-1} \epsilon^2 L_T \omega(\gamma^{-2} L_T). \end{aligned} \quad (5.39)$$

Then use (5.39), (5.38) in order to conclude the proof. □

## 6 Ordered processes

The control of processes like, for instance:

$$\eta(t) = \sum_{k=1}^{+\infty} \epsilon^2 \lambda_k^2(t) b_k^{-2} (\xi_k^2 - 1), \quad t \geq 0 \text{ or } \rho(t) = \sum_{k=1}^{+\infty} (1 - \lambda_k(t)) \theta_k \epsilon b_k^{-1} \xi_k, \quad t \geq 0, \quad (6.1)$$

is rather important for the construction of a data-driven parameter choice rule. These processes have a strong influence on the behavior of the related estimators. For all  $t > 0$ , the sequence  $(\lambda_k(t))_{k \in \mathbb{N}}$  is assumed to be monotone decreasing.

We will see in this section that they are embedded in a more general class: the ordered processes. These stochastic objects have been introduced in [20] and are intensively studied in [7]. In this section we recall the definition and present some results useful for the proofs of Theorems 1-3.

**Definition 2** Let  $\zeta(t)$ ,  $t \geq 0$  a separable random process with  $\mathbb{E}\zeta(t) = 0$  and finite variance  $\Sigma^2(t)$ . It is called ordered if for all  $t_2 \geq t_1 \geq 0$ ,

$$\Sigma^2(t_2) \geq \Sigma^2(t_1), \text{ and } \mathbb{E}[\zeta(t_2) - \zeta(t_1)]^2 \leq \Sigma^2(t_2) - \Sigma^2(t_1). \quad (6.2)$$

The class of ordered processes is thus rather vast. In particular, it contains the well-known Wiener processes.

**Proposition 1** The process  $\eta(t)$ ,  $t \geq 0$ , defined in (6.1) is ordered.

PROOF. For all  $t \geq 0$ ,

$$\mathbb{E}\eta(t) = 0 \text{ and } \Sigma^2(t) = 2\epsilon^4 \sum_{k=1}^{+\infty} \lambda_k^4(t) b_k^{-4} = 2\sigma^2(t).$$

Assume that for all  $k \in \mathbb{N}$ ,  $\lambda_k(t) \rightarrow 1$  as  $t \rightarrow +\infty$  and let  $t_1 \leq t_2$  be fixed. Then  $\Sigma^2(t_1) \leq \Sigma^2(t_2)$  since, for all  $k \in \mathbb{N}$ ,  $\lambda_k(t_1) \leq \lambda_k(t_2)$ . Moreover,

$$\begin{aligned} \mathbb{E}\eta(t_1)^2 &= \mathbb{E} \left[ \sum_{k=1}^{+\infty} \epsilon^2 \lambda_k^2(t_1) b_k^{-2} (\xi_k^2 - 1) \right]^2 = \epsilon^4 \sum_{k=1}^{+\infty} \lambda_k^4(t_1) b_k^{-4} \mathbb{E} [(\xi_k^2 - 1)^2], \\ &\leq \sum_{k=1}^{+\infty} \epsilon^2 \lambda_k^2(t_1) b_k^{-2} \epsilon^2 \lambda_k^2(t_2) b_k^{-2} \mathbb{E} [(\xi_k^2 - 1)^2], \\ &= \mathbb{E} \left[ \left( \sum_{k=1}^{+\infty} \epsilon^2 \lambda_k^2(t_1) b_k^{-2} (\xi_k^2 - 1) \right) \times \left( \sum_{k=1}^{+\infty} \epsilon^2 \lambda_k^2(t_2) b_k^{-2} (\xi_k^2 - 1) \right) \right] = \mathbb{E}[\eta(t_1)\eta(t_2)] \end{aligned} \quad (6.3)$$

The property (6.2) is a consequence of (6.3). The process  $\eta$  is ordered. □

The main characteristic of ordered processes is contained in the following lemma.

**Lemma 2** *Let  $\zeta(t)$ ,  $t \geq 0$  an ordered process and suppose that there exists  $\kappa > 0$  such that*

$$\varphi(\kappa) = \sup_{t_1, t_2} \mathbb{E} \exp \left\{ \kappa \frac{\zeta(t_1) - \zeta(t_2)}{\sqrt{\mathbb{E}[\zeta(t_1) - \zeta(t_2)]^2}} \right\} < +\infty. \quad (6.4)$$

*Then, there exists a constant  $C$  depending on  $\kappa$  such that, for all  $S, T > 0$  and all  $p \geq 1$ ,*

$$\mathbb{E} \sup_{s, t \in [S, T]} |\zeta(t) - \zeta(s)|^p \leq C^p p^p \sqrt{\Sigma^2(T) - \Sigma^2(S)}^p. \quad (6.5)$$

This result is an extension of Lemma 1 of [7]. They consider the particular case  $S = 0$ . This lemma is rather important. It may be useful in many situations. In particular, many processes encountered in the proofs of Theorems 1 and 2 are ordered and satisfy (6.4). The proof follows essentially the same lines of [7].

The following result is a consequence of the previous lemma.

**Lemma 3** *Let  $\zeta(t)$ ,  $t \geq 0$  be an ordered process satisfying (6.4) such that  $\zeta(0) = 0$  and  $t^*$  measurable with respect to  $\zeta$ . Then there exists a positive constant  $C$  depending on  $\kappa$  and  $\tau > 0$  such that for all  $K > 1$ :*

$$\mathbb{E}|\zeta(t^*)| \leq \log^2(K) \mathbb{E}\Sigma(t^*) + C \frac{\sqrt{\mathbb{E}\Sigma^2(t^*)}}{K}.$$

PROOF. First remark that:

$$\begin{aligned} \mathbb{E}\zeta(t^*) &= \mathbb{E}\zeta(t^*) \mathbf{1}_{\{\zeta(t^*) \leq (\log K)^2 \Sigma(t^*)\}} + \mathbb{E}\zeta(t^*) \mathbf{1}_{\{\zeta(t^*) \geq (\log K)^2 \Sigma(t^*)\}}, \\ &\leq (\log K)^2 \mathbb{E}\Sigma(t^*) + \mathbb{E}\zeta(t^*) \mathbf{1}_{\{\zeta(t^*) \geq (\log K)^2 \Sigma(t^*)\}}. \end{aligned} \quad (6.6)$$

We can assume without loss of generality that  $\Sigma(t) \rightarrow +\infty$  as  $t \rightarrow +\infty$ . Let  $(t_k)_{k \in \mathbb{N}}$  the real sequence verifying:

$$\Sigma(t_k) = k^d \frac{\sqrt{\mathbb{E}\Sigma^2(t^*)}}{K}, \quad \forall k \in \mathbb{N}, \quad (6.7)$$

where  $d > 0$  will be chosen later. Using Lemma 2,

$$\begin{aligned} &\mathbb{E}\zeta(t^*) \mathbf{1}_{\{\zeta(t^*) \geq \log^2(K) \Sigma(t^*)\}} \\ &= \mathbb{E}\zeta(t^*) \mathbf{1}_{\{\zeta(t^*) \geq \log^2(K) \Sigma(t^*)\}} \mathbf{1}_{\{t^* < t_1\}} + \mathbb{E}\zeta(t^*) \mathbf{1}_{\{\zeta(t^*) \geq \log^2(K) \Sigma(t^*)\}} \mathbf{1}_{\{t^* > t_1\}}, \\ &\leq \mathbb{E} \sup_{t < t_1} |\zeta(t)| + \mathbb{E}\zeta(t^*) \mathbf{1}_{\{\zeta(t^*) \geq \log^2(K) \Sigma(t^*)\}} \mathbf{1}_{\{t^* > t_1\}}, \\ &\leq \Sigma(t_1) + \mathbb{E}\zeta(t^*) \mathbf{1}_{\{\zeta(t^*) \geq \log^2(K) \Sigma(t^*)\}} \mathbf{1}_{\{t^* > t_1\}}, \\ &= \frac{\sqrt{\mathbb{E}\Sigma^2(t^*)}}{K} + \mathbb{E}\zeta(t^*) \mathbf{1}_{\{\zeta(t^*) \geq \log^2(K) \Sigma(t^*)\}} \mathbf{1}_{\{t^* > t_1\}}. \end{aligned}$$

Let  $p \in \mathbb{N}$  such that  $1 < p < 2$ . Using Hölder inequality:

$$\begin{aligned}
& \mathbb{E} \zeta(t^*) \mathbf{1}_{\{\zeta(t^*) \geq \log^2(K) \Sigma(t^*)\}} \mathbf{1}_{\{t^* > t_1\}} \\
&= \sum_{k=1}^{+\infty} \mathbb{E} \zeta(t^*) \mathbf{1}_{\{\zeta(t^*) \geq \log^2(K) \Sigma(t^*)\}} \mathbf{1}_{\{t \in [t_k, t_{k+1}]\}}, \\
&= \sum_{k=1}^{+\infty} \mathbb{E} \left[ \Sigma^p(t^*) \frac{\zeta(t^*)}{\Sigma^p(t^*)} \mathbf{1}_{\{\zeta(t^*) \geq \log^2(K) \Sigma(t^*)\}} \mathbf{1}_{\{t \in [t_k, t_{k+1}]\}} \right], \\
&\leq \sum_{k=1}^{+\infty} [\mathbb{E} \Sigma^{pr}(t^*)]^{1/r} \times \left[ \mathbb{E} \frac{|\zeta(t^*)|^s}{\Sigma^{ps}(t^*)} \mathbf{1}_{\{\zeta(t^*) \geq \log^2(K) \Sigma(t^*)\}} \mathbf{1}_{\{t \in [t_k, t_{k+1}]\}} \right]^{1/s},
\end{aligned}$$

where  $r$  and  $s$  are such that  $r^{-1} + s^{-1} = 1$ . In the following, set  $r = 2/p$  and  $s = 2/(2-p)$ . Such a choice leads to:

$$\begin{aligned}
& \mathbb{E} \zeta(t^*) \mathbf{1}_{\{\zeta(t^*) \geq \log^2(K) \Sigma(t^*)\}} \mathbf{1}_{\{t^* > t_1\}} \\
&\leq \sqrt{\mathbb{E} \Sigma^2(t^*)}^p \times \sum_{k=1}^{+\infty} \left[ \frac{\mathbb{E} \sup_{t \in [t_k, t_{k+1}]} |\zeta(t)|^s}{\Sigma^{ps}(t_k)} \mathbf{1}_{\{\sup_{t \in [t_k, t_{k+1}]} |\zeta(t)| \geq \log^2(K) \Sigma(t_k)\}} \right]^{1/s}.
\end{aligned}$$

Let  $q > 0$  which will be chosen later. Using a Markov inequality and Lemma 2, we obtain:

$$\begin{aligned}
& \mathbb{E} \zeta(t^*) \mathbf{1}_{\{\zeta(t^*) \geq (\log K) \Sigma(t^*)\}} \mathbf{1}_{\{t^* > t_1\}} \\
&\leq \sqrt{\mathbb{E} \Sigma^2(t^*)}^p \times \sum_{k=1}^{+\infty} \left[ \frac{\mathbb{E} \sup_{t \in [t_k, t_{k+1}]} |\zeta(t)|^{s+q}}{\Sigma^{ps+q}(t_k)} \times \frac{1}{(\log K)^{2q}} \right]^{1/s}, \\
&\leq \sqrt{\mathbb{E} \Sigma^2(t^*)}^p \times \sum_{k=1}^{+\infty} \left[ \frac{C(s+q)^{(s+q)} \Sigma^{s+q}(t_{k+1})}{\Sigma^{ps+q}(t_k)} \times \frac{1}{(\log K)^{2q}} \right]^{1/s}.
\end{aligned}$$

Now, use (6.7) in order to obtain:

$$\begin{aligned}
& \mathbb{E} \zeta(t^*) \mathbf{1}_{\{\zeta(t^*) \geq \log^2 K \Sigma(t^*)\}} \mathbf{1}_{\{t^* > t_1\}} \\
&\leq \sqrt{\mathbb{E} \Sigma^2(t^*)}^p \times \sum_{k=1}^{+\infty} \left[ \frac{C(s+q)^{(s+q)} (k+1)^{d(s+q)} \sqrt{\mathbb{E} \Sigma^2(t^*)}^{s+q} K^{-(s+q)}}{k^{d(ps+q)} \sqrt{\mathbb{E} \Sigma^2(t^*)}^{ps+q} K^{-(ps+q)}} \times \frac{1}{(\log K)^{2q}} \right]^{1/s}, \\
&\leq \sqrt{\mathbb{E} \Sigma^2(t^*)}^p \times \sum_{k=1}^{+\infty} \left[ \frac{C(s+q)^{(s+q)} 2^{d(s+q)} K^{s(p-1)}}{k^{ds(p-1)} \sqrt{\mathbb{E} \Sigma^2(t^*)}^{s(p-1)}} \times \frac{1}{(\log K)^{2q}} \right]^{1/s}, \\
&= \frac{\sqrt{\mathbb{E} \Sigma^2(t^*)}^p}{\sqrt{\mathbb{E} \Sigma^2(t^*)}^{p-1}} \times \frac{1}{K} \times \sum_{k=1}^{+\infty} \frac{1}{k^{d(p-1)}} \left[ \frac{C(s+q)^{(s+q)} 2^{d(s+q)} K^{sp}}{(\log K)^{2q}} \right]^{1/s}.
\end{aligned}$$

Setting for instance  $d = 2/(p-1)$ ,

$$\mathbb{E} \zeta(t^*) \mathbf{1}_{\{\zeta(t^*) \geq \log^2(K) \Sigma(t^*)\}} \mathbf{1}_{\{t^* > t_1\}} \leq C \frac{\sqrt{\mathbb{E} \Sigma^2(t^*)}}{K} \times \frac{q^{q/s} 2^{dq} K^p}{(\log K)^{2q/s}}.$$

Recall that the parameter  $1 < p < 2$  is fixed. Set  $q = s \log K$ . Therefore:

$$\begin{aligned}
& \mathbb{E} \zeta(t^*) \mathbf{1}_{\{\zeta(t^*) \geq \log^2(K) \Sigma(t^*)\}} \mathbf{1}_{\{t^* > t_1\}} \\
&\leq C \frac{\sqrt{\mathbb{E} \Sigma^2(t^*)}}{K} \times \frac{K^p (\log K)^{\log K} 2^{d \log K/s}}{(\log K)^{2 \log K}}, \\
&= C \frac{\sqrt{\mathbb{E} \Sigma^2(t^*)}}{K} \times \left( \frac{e^p 2^d}{\log K} \right)^{\log K} \leq C \frac{\sqrt{\mathbb{E} \Sigma^2(t^*)}}{K},
\end{aligned}$$

for some  $C > 0$  independent of  $K$ . This concludes the proof of Lemma 3.

□

**Corollary 1** *Let  $\hat{t} > 0$  measurable w.r.t. the sequence  $(\xi_k)_{k \in \mathbb{N}}$  and  $t^0$  the oracle defined in (3.5). For all  $\gamma > 0$ , the following inequalities hold:*

$$\begin{aligned} \text{(i)} \quad & \mathbb{E}_\theta \left| \sum_{k=1}^{+\infty} \lambda_k^2(\hat{t}) \theta_k \epsilon b_k^{-1} \xi_k \right| \leq \gamma \mathbb{E}_\theta R_\epsilon(\theta, \hat{\lambda}) + \gamma R_\epsilon(\theta, \lambda_0) + C \epsilon^2 \gamma^{-1} L_T \omega(\gamma^{-2} L_T), \\ \text{(ii)} \quad & \mathbb{E}_\theta \left| \sum_{k=1}^{+\infty} (1 - \lambda_k(\hat{t}))^2 \theta_k \epsilon b_k^{-1} \xi_k \right| \leq \gamma \mathbb{E}_\theta R_\epsilon(\theta, \hat{\lambda}) + \gamma R_\epsilon(\theta, \lambda_0) + C \epsilon^2 \gamma^{-1} L_T \omega(\gamma^{-2} L_T), \end{aligned}$$

for some  $C > 0$  independent of  $\epsilon$ .

PROOF. In a first time introduce:

$$\Delta^\epsilon[\lambda] = \epsilon^2 \sup_{k \in \mathbb{N}} \lambda_k^2 b_k^{-2}, \quad \forall \lambda \in \Lambda. \quad (6.8)$$

First prove inequality (i).

$$\mathbb{E}_\theta \sum_{k=1}^{+\infty} \lambda_k^2(\hat{t}) \theta_k \epsilon b_k^{-1} \xi_k = \mathbb{E}_\theta \sum_{k=1}^{+\infty} \{\lambda_k^2(\hat{t}) - \lambda_k^2(t_0)\} \theta_k \epsilon b_k^{-1} \xi_k.$$

The process:

$$\rho(t) = \sum_{k=1}^{+\infty} \{\lambda_k^2(t) - \lambda_k^2(t_0)\} \theta_k \epsilon b_k^{-1} \xi_k, \quad t > 0,$$

is ordered on  $[t_0; +\infty[$ . Moreover,  $\tilde{\rho}(t) = \rho(t^{-1})$ ,  $t \geq 0$  is also ordered on  $[t_0^{-1}; +\infty[$ . Both processes satisfy (6.4) for some  $\kappa > 0$ . Hence, using Lemma 3 with  $K = \sqrt{S}$ :

$$\begin{aligned} & \mathbb{E}_\theta \sum_{k=1}^{+\infty} \lambda_k^2(\hat{t}) \theta_k \epsilon b_k^{-1} \xi_k \\ & \leq C \log^2(S) \cdot \mathbb{E}_\theta \sqrt{\sum_{k=1}^{+\infty} \{\lambda_k^2(t_0) - \lambda_k^2(\hat{t})\}^2 \theta_k^2 \epsilon^2 b_k^{-2}} + \frac{C}{\sqrt{S}} \sqrt{\mathbb{E}_\theta \sum_{k=1}^{+\infty} \{\lambda_k^2(t_0) - \lambda_k^2(\hat{t})\}^2 \theta_k^2 \epsilon^2 b_k^{-2}}, \end{aligned}$$

where  $\lambda^0 = \lambda(t_0)$  is the oracle defined in (3.5) and  $C$  a positive constant independent of  $\epsilon$ . Then, remark that the following inequality holds:

$$\begin{aligned} \{\lambda_k^2(t_0) - \lambda_k^2(\hat{t})\}^2 &= [(1 - \lambda_k(\hat{t})) - (1 - \lambda_k(t_0))]^2 (\lambda_k(\hat{t}) + \lambda_k(t_0))^2, \\ &\leq 4 \left[ (1 - \lambda_k(\hat{t}))^2 + (1 - \lambda_k(t_0))^2 \right] [\lambda_k^2(t_0) + \lambda_k^2(\hat{t})]. \end{aligned}$$

For all  $\gamma > 0$ , using the elementary inequality  $2ab \leq \gamma a^2 + \gamma^{-1} b^2$ , we obtain:

$$\begin{aligned} & \mathbb{E}_\theta \sum_{k=1}^{+\infty} \lambda_k^2(\hat{t}) \theta_k \epsilon b_k^{-1} \xi_k \\ & \leq \gamma \mathbb{E}_\theta \sum_{k=1}^{+\infty} (1 - \hat{\lambda}_k)^2 \theta_k^2 + D \gamma^{-1} \log^2(S) \cdot \mathbb{E}_\theta \Delta^\epsilon[\hat{\lambda}] + \gamma \sum_{k=1}^{+\infty} (1 - \lambda_k^0)^2 \theta_k^2 + D \gamma^{-1} \log^2(S) \cdot \Delta^\epsilon[\lambda^0], \end{aligned}$$

for some positive constant  $D$ . With simple algebra, for all  $x > 0$  and  $\lambda \in \Lambda$ :

$$\begin{aligned} D \gamma^{-1} \log^2(S) \cdot \Delta^\epsilon[\lambda] &= D \gamma^{-1} \log^2(S) \cdot \epsilon^2 \sup_k \lambda_k^2 b_k^{-2}, \\ &\leq D \gamma^{-1} \log^2(S) \cdot x^{-1} \epsilon^2 \sum_{k=1}^{+\infty} \lambda_k b_k^{-2} + D \gamma^{-1} \log^2(S) \cdot \epsilon^2 \omega(x), \end{aligned} \quad (6.9)$$

where the function  $\omega$  is introduced in (3.6) and computed for some particular examples in Section 3.2. Then set  $x = D \gamma^{-2} \log^2 S$  in order to obtain (i).

The proof of (ii) exactly follows the same lines. It uses in particular Corollary 1 and inequality (34) of [8].

□

## References

- [1] A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Probability Theory and Related Fields*, 113:301–413, 1999.
- [2] F. Bauer and T. Hohage. A Lepskij-type stopping rule for regularized Newton methods. *Inverse Problems*, 21:1975–1991, 2005.
- [3] L. Birgé and P. Massart. Gaussian model selection. *J. Eur. Math. Soc.*, 3:203–268, 2001.
- [4] N. Bissantz, G. Claeskens, H. Holzmann, and A. Munk. Testing for lack of fit in inverse regression – with applications to biophotonic imaging. *J. Royal Statist. Soc. Ser. B*, 71 (1):25–48, 2009.
- [5] N. Bissantz, T. Hohage, and A. Munk. Consistency and rates of convergence of nonlinear Tikhonov regularization with random noise. *Inverse Problems*, 20:1773–1789, 2004.
- [6] N. Bissantz, T. Hohage, A. Munk, and F. Ruyngaert. Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM J. Numerical Analysis*, 45:2610–2636, 2007.
- [7] Y. Cao and Y. Golubev. On oracle inequalities related to smoothing splines. *Math. Methods Statist.*, 15:398–414, 2006.
- [8] L. Cavalier, G.K. Golubev, D. Picard, and A.B. Tsybakov. Oracle inequalities for inverse problems. *Annals of Statistics*, 30:843–874, 2002.
- [9] L. Cavalier and Y. Golubev. Risk hull method and regularization by projections of ill-posed inverse problems. *Annals of Statistics*, 34:1653–1677, 2006.
- [10] L. Cavalier and A.B. Tsybakov. Sharp adaptation for inverse problems with random noise. *Probability Theory and Related Fields*, 123:323–354, 2002.
- [11] D.L. Donoho. Nonlinear solutions of linear inverse problems by wavelet-vaguelette decomposition. *Journal of Applied and Computational Harmonic Analysis*, 2:101–126, 1995.
- [12] S. Efromovich. Robust and efficient recovery of a signal passed through a filter and then contaminated by non-gaussian noise. *IEEE Trans. Inform. Theory*, 43:1184–1191, 1997.
- [13] H.W. Engl. On the choice of the regularization parameter for iterated Tikhonov regularization of ill-posed problems. *Journal of approximation theory*, 49:55–63, 1987.
- [14] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers Group, Dordrecht, 1996.
- [15] M.S. Ermakov. Minimax estimation of the solution of an ill-posed convolution type problem. *Problems of Information Transmission*, 25:191–200, 1989.
- [16] Yu. Golubev. The principle of penalized empirical risk in severely ill-posed problems. *Theory Probab. and Appl.*, 130:18–38, 2004.
- [17] M. Hanke. Accelerated Landweber iterations for the solution of ill-posed equations. *Numerische mathematik*, 60:341–373, 1991.
- [18] T. Hida. *Brownian Motion*. Springer-Verlag, New York-Berlin, 1980.
- [19] I.M. Johnstone and B.W. Silverman. Speed of estimation in positron emission tomography and related inverse problems. *Annals of Statistics*, 18:251–280, 1990.
- [20] A. Kneip. Ordered linear smoother. *Annals of Statistics*, 22:835–866, 1994.
- [21] J.M. Loubes and C. Ludena. Penalized estimators for non-linear inverse problems. 2007. To appear in ESAIM Probability and Statistics.



- [22] C. Marteau. On the stability of the risk hull method for projection estimator. 2007. To appear in Journal of Statistical Planning and Inference.
- [23] P. Mathé. The Lepskij principle revisited. *Inverse problems*, 22:L11–L15, 2006.
- [24] P. Mathé and S.V. Pereverzev. Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projection methods. *SIAM J. Numer. Anal.*, 38:1999–2021, 2001.
- [25] D.N. Ghosh Roy and L.S. Couchman. *Inverse problems and inverse scattering of plane waves*. Academic press San Diego, 2002.
- [26] C.M. Stein. Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, 9:1135–1151, 1981.